SAMPLING PROPERTIES OF SOME ECONOMETRIC TESTS IN THE PRESENCE OF MODEL MIS-SPECIFICATION

A thesis submitted pursuant to the requirements for the award of the Doctor of Philosophy degree at the University of Canterbury

by

John P. Small.

April 1993

TABLE OF CONTENTS

Acknowledgements....1

1 Introduction....3

2 Hypothesis Testing....11

2.1 Motivation....11

2.2 Foundations....15

2.3 LR LM and Wald Tests....24

2.4 Exact Tests....31

3 Testing Regression Errors for the Presence of a Scalar Covariance Matrix....34

3.1 Tests of Serial Independence....35

3.2 Heteroscedasticity Tests....47

3.2.1 Unconditional Heteroscedasticity....48

3.2.2 Conditional Heteroscedasticity....52

3.3 Conclusion....56

4 Testing for Serial Independence in Mis-specified Linear Regression Models....57

4.1 Non-Normal Errors....58

4.2 Omitted Variables....61

4.3 Stochastic Regressors....63

4.4 Incorrect Alternatives....66

4.5 Heteroscedasticity....................................70

5 The Simultaneous Modelling of Autocorrelation and Heteroscedasticity....................................79

5.1 Unconditional Heteroscedasticity..................81

5.2 Conditional Heteroscedasticity....................87

5.3 Conclusion....................................95

Appendix 5.1....................................96

6 Testing Against AR(1) Alternatives in the Presence of Unconditional Heteroscedasticity....................99

6.1 Introduction....................................99

6.2 Model and Tests....................................100

6.3 Theoretical Discussion....................................105

6.4 The Data....................................115

6.5 Numerical Results....................................117

6.5.1 Results Using V*....................................117

6.5.2 Results Using V**....................................121

6.6 Conclusion....................................124

Tables and Graphs....................................125

7 Testing for Serial Independence Against Incorrect Alternatives....................................139

7.1 Introduction....................................139

7.2 Testing Against AR(1) in the Presence of AR(2) Errors....................................142

7.2.1 Theoretical Issues....................................142
7.2.2 Numerical Evaluations....................................146
7.2.3 Conclusion....................................151
Tables and Graphs....................................152
7.3 Testing Against AR(1) in the Presence of MA(1) Errors....................................159
7.3.1 Theoretical Issues....................................160
7.3.2 Numerical Evaluations....................................162
7.3.3 Conclusion....................................165
Tables and Graphs....................................166
7.4 Testing With Seasonal Autoregressive Mis-Specification....................................173
7.4.1 Theoretical Issues....................................174
7.4.2 Numerical Evaluations....................................178
7.4.3 Conclusion....................................180
Tables and Graphs....................................181
7.5 Conclusion....................................188

8 Testing the Conditional Mean Under Conditional Variance Mis-Specification....................................191
8.1 Introduction....................................191
8.2 The Models....................................193
8.3 The Tests....................................195
8.4 Experimental Design....................................202
8.5 Experimental Results....................................204
8.5.1 Basic Properties....................................205

8.5.2 Sample Size Effects....................211
8.5.3 Relaxing Conditional Normality..................214
8.6 Conclusion....................216
Tables and Graphs....................218
9 Conclusion....................237
References....................242

LIST OF TABLES

Table Number	Description	Page
6.1	Power of DW and s(0.5) tests, multiplicative heteroscedasticity, Spirits data	125
6.2	Power of ADW and BW tests, multiplicative heteroscedasticity, CPI data	126
6.3	Power of DW and s(0.5) tests, additive heteroscedasticity, Normal data	127
6.4	Power of DW and s(0.75) tests, additive heteroscedasticity, Uniform data	128
6.5	Power of ADW and BW tests, "Chow-type" heteroscedasticity, Lognormal data	129
6.6	Power of ADW and BW tests, "Chow-type" heteroscedasticity, Watson's data	130
6.7	Selected powers, multiplicative heteroscedasticity assuming V** covariance matrix	131
6.8	Selected powers, multiplicative heteroscedasticity assuming V** covariance matrix	132
7.2.1	Power of ADW test with AR(2) errors, CPI data	152
7.2.2	Power of ADW test with AR(2) errors, CPI data	153
7.2.3	Power of BW test with AR(2) errors, Unemployment data	154
7.2.4	Power of BW test with AR(2) errors, Unemployment data	155
7.3.1	Power of DW and BW tests with MA(1), AR(1) errors, Normal data	166
7.3.2	Power of DW and BW tests with MA(1), AR(1) errors, Uniform data	167
7.3.3	Power of DW and BW tests with MA(1), AR(1) errors, Lognormal data	168
7.3.4	Power of DW and BW tests with MA(1), AR(1) errors, Watson's data	169

Table Number	Description	Page
7.4.1	Power of ADW and s(0.75) tests with restricted AR(5) errors, CPI data	181
7.4.2	Power of DW and s(0.5) tests with restricted AR(5) errors, Lognormal data	182
7.4.3	Power of DW and BW tests with restricted AR(5) errors, Uniform data	183
7.4.4	Power of ADW and BW tests with restricted AR(5) errors, Watson's data	184
8.1	Power of $s_1(0.5)$ test with ARCH errors, X8 data	218
8.2	Power of DW_4 test with ARCH errors, X8 data	219
8.3	Power of BP and LB tests with ARCH errors, X9 data	220
8.4	Power of LB and LBA tests with ARCH errors, X9 data	221
8.5	Power of LM and WLM tests with ARCH errors, X9 data	222
8.6	Power of BP and BPA tests with GARCH errors, X11 data	223
8.7	Power of LM test with GARCH errors, X11 data, Normal and Student-t_4 distributed innovations	224

LIST OF FIGURES

Figure Number	Description	Page
6.1a	Power of several tests, no heteroscedasticity, Uniform data	133
6.1b	Power of several tests, no heteroscedasticity, Watson's data	133
6.2a	Power of several tests, "Chow-type" heteroscedasticity, Uniform data	134
6.2b	Power of several tests, "Chow-type" heteroscedasticity, Watson's data	134
6.3a	Power of DW test, additive heteroscedasticity, CPI data, T=60	135
6.3b	Power of DW test, additive heteroscedasticity, CPI data, T=20	135
6.4a	Power of DW test, "Chow-type" heteroscedasticity, Spirits data	136
6.4b	Power of s(0.5) test, additive heteroscedasticity, Uniform data	136
6.5a	Power of several tests, no heteroscedasticity, Normal data, V** covariance matrix	137
6.5b	Power of several tests, multiplicative heteroscedasticity, Normal data, V** covariance matrix	137
6.6a	Power of DW test, "Chow-type" heteroscedasticity, Spirits data, V** covariance matrix	138
6.6b	Power of ADW test, additive heteroscedasticity, Uniform data, V** covariance matrix	138
7.2.1	Stationarity region for an AR(2) process	144
7.2.2	Power of ADW test, AR(2) errors, Spirits data	156
7.2.3	Power of BW test, AR(2) errors, Watson's data	156
7.2.4	Power of DW test, AR(2) errors, Unemployment data	157
7.2.5	Power of s(0.5) test, AR(2) errors, Watson's data	157

7.2.6 Power of ADW test, AR(2) errors, Watson's data........158

7.2.7 Power of DW test, AR(2) errors, Watson's data.........158

7.3.1 Power of ADW test, MA(1) errors, Unemployment data....170

7.3.2 Power of BW test, MA(1) errors, Uniform data..........170

7.3.3 Power of several tests, MA(1) errors, CPI data........171

7.3.4 Power of DW test, MA(1) errors, Watson's data.........171

7.3.5 Power of s(0.5) test, MA(1) errors, Watson's data.....172

7.3.6 Power of BW and DW tests, MA(1) errors, Watson's data...173

7.4.1 Power of several tests, AR(5) errors, ϕ_4=-0.8, Spirits data...185

7.4.2 Power of several tests, AR(5) errors, ϕ_4=0.8, Spirits data...185

7.4.3 Power of several tests, AR(5) errors, ϕ_4=0.6, Unemployment data....................................186

7.4.4 Power of several tests, AR(5) errors, ϕ_4=0.8, Unemployment data....................................186

7.4.5 Power of several tests, AR(5) errors, ϕ_4=-1, Normal data..187

7.4.6 Power of several tests, AR(5) errors, ϕ_1=0.8, Spirits data...187

8.1a Power of s(0.75) test, ARCH errors, X10 data, Weiss model..225

8.1b Power of s(0.75) test, GARCH errors, X10 data, BHL model..225

8.2a Power of BP and LB tests, ARCH errors, X9 data, Weiss model..226

8.2b Power of BP and LB tests, GARCH errors, X9 data, BHL model..226

8.3a Power of LBA test, ARCH errors, X11 data..............227

8.3b Power of LBA test, GARCH errors, X11 data.............227

8.4a Power of several tests, ARCH errors, X8 data, Weiss Model....228

8.4b Power of several tests, GARCH errors, X8 data, BHL Model....228

8.5a Power of LM test, ARCH errors, X8 data, BHL model....229

8.5b Power of LM test, ARCH errors, X8 data, Weiss model...229

8.6a Power of LM and WLM tests, ARCH errors, X10 data, Weiss model....230

8.6b Power of LM and WLM tests, GARCH errors, X6 data, BHL model....230

8.7a Power of DW test, Student-t_4 ARCH errors, X8 data....231

8.7b Power of LB and LBA tests, Student-t_4 GARCH errors, X11 data....231

8.8a Power of LM test, Student-t_4 ARCH errors, X8 data....232

8.8b Power of LM and WLM tests, Student-t_4 errors, X8 data....232

8.9a Power of several tests, ARCH errors, X6 data, Weiss model....233

8.9b Power of several tests, GARCH errors, X6 data, BHL model....233

8.10a Power of several tests, ARCH errors, X9 data, Weiss model....234

8.10b Power of several tests, GARCH errors, X9 data, BHL model....234

8.11a Power of several tests, ARCH errors, X10 data, Weiss model....235

8.11b Power of several tests, GARCH errors, X10 data, BHL model....235

8.12a Power of several tests, ARCH errors, X11 data, Weiss model....236

8.12b Power of several tests, GARCH errors, X11 data, BHL model....236

ACKNOWLEDGEMENTS

The completion of this thesis is the culmination of my years of formal study in econometrics, which began early in 1989. At various times during this period David Giles has been my teacher, mentor, inspiration, supervisor and friend. Without his presence this thesis would certainly not have been written and I consider myself extremely fortunate to have been associated with David over the last four years. He has encouraged me to set rigorous standards, to tackle interesting and relevant research problems and to present my work to the academic community on a regular basis. In short, he has shown me how to be an effective scholar.

I would also like to acknowledge the important role of my parents, John and Dell, who showed me the value of an academic career, and then allowed me the freedom to follow my own interests from an early age.

Several other colleagues in the Department of Economics at the University of Canterbury deserve special mention. The excellent research produced by Judith Giles has been an inspiration to me and her comments on early sections of this work greatly assisted their clarity. Robin Harrison and Alan Wan have also made useful comments at various stages. Members of the microeconomics group have continued to stimulate my interest in economics generally which has been helpful during the times when my research has appeared less than vital. In this regard I thank Michael Carter, John Fountain, Steve Shea and Alan Woodfield.

At various times specialists from outside this institution have read and commented on sections of this work. This is a relatively thankless task which is very much appreciated. I wish to thank Robert Bartels, Howard Doran, Max King and Mike Veall, all of whom have all contributed to this informal reviewing process.

Financial support has been provided by a New Zealand Vice-Chancellors Committee Scholarship and part-time teaching in the Department of Economics. The assistance of the Department of Economics went further than this however and I thank Alan Woodfield as Head of Department in particular, for his support of my conference expenses and journal submission fees.

Finally, but most importantly, I thank my wife Lynne without whose love and support I could not have completed this work. She raised my spirits and kept me on track during those occasions when I questioned the value of my endeavours. Lynne was a full time student herself during the first year of this research but could still find the energy to support me when it was necessary. Our children, Katie and Peter, also deserve considerable credit for their tolerance of the stresses associated with having two students for parents.

CHAPTER 1

INTRODUCTION

The objective of this thesis is to discover the consequences of applying certain diagnostic tests relating to the errors of a regression model which is mis-specified in one of several different ways. All of the tests considered are based on the null hypothesis of serial independence and all are in regular use by applied econometricians. Similarly, the non-standard conditions under which these tests are analysed are intended to represent situations which can reasonably be expected to occur in practice. It is therefore expected that the results reported in this thesis will be of direct benefit to the econometrics profession in the form of an improved understanding of the effects of mis-specification.

Why should we be interested in the consequences of model mis-specification in econometrics? The answer to this question provides the fundamental motivation for the original analysis contained in this thesis.

An econometric model provides the means for addressing economic questions using appropriate statistical procedures. The non-experimental nature of most economic data means that extra care must be taken to validate the chosen model. In particular, the assumptions underlying the random components of the model must be verified carefully. It is important to establish, for example, the characteristics of the error variance, so that the appropriate technique can be employed to estimate the parameters of the model. If the error variance is incorrectly assumed to be constant, then Ordinary Least Squares (OLS) estimators of the parameters which characterise the mean of the dependent variable are inefficient, although they remain unbiased. More seriously, the OLS estimator of the error variance in this case is biased (downwards, in a simple regression involving two positively correlated variables), leading to invalid "t" statistics.

Careful scrutiny of the underlying statistical assumptions is the foundation of good econometric practice, as the above example illustrates. A great variety of tests, which can be used to assist this examination, have been published in the econometrics literature. The power[1] properties of these tests are, however, very diverse and strongly dependent on the conditions under which they are applied. In general, the publication of a new test is accompanied by evidence of its power against a specific departure from the null hypothesis. When this state corresponds to the alternative hypothesis which the test is designed to detect, we say that the test has been applied in a correctly specified model. High power under correct specification is a basic requirement for the general acceptance of a test.

From the viewpoint of the applied researcher, however,

[1] Formally, the power of a test is the probability that it will reject the null hypothesis, conditional on the state of nature.

"optimal" power in a correctly specified model ***should be*** a necessary, but not a sufficient condition for the widespread adoption of a test. The reason for this is simple. In any given model there are several violations of maintained statistical assumptions which are possible and should be tested for. In many cases, econometric tests themselves assume that only one problem exists: that which the test is designed to detect. Thus, an ostensibly rigorous series of diagnostic tests, intended to test the validity of the maintained assumptions, may itself be invalid due to the violation of maintained assumptions. We need, therefore, to know how tests are affected by the mis-specification of the model, that is, by the presence of conditions which are different from those on which the test is based.

The emphasis of the previous paragraph is on the normative rather than the positive aspects of econometric hypothesis testing. This reflects the current state of knowledge about the effects of mis-specification. It is simply not possible to confine diagnostic testing to procedures which are known to be robust to the violation of other assumptions because insufficient research has been directed towards investigating such issues. The original work reported in this thesis is a contribution towards the knowledge which is needed to resolve the logical inconsistencies which are inherent in most programs of diagnostic testing.

Several interesting aspects of this overall objective are not dealt with here. We do not, for example, consider Bayesian approaches to the problem of detecting autocorrelation in mis-specified models. Neither is there a temporal aspect to our analysis such as would be found in a pre-test study. The final major abstraction is from nonlinear regression models *per se*. The conditional variance models used in chapter 8 are nonlinear but these are considered in the context of a standard linear regression model for the mean of the dependent variable.

The next chapter begins with a discussion of the motivations for testing hypotheses in econometrics. This is followed by an analysis of the statistical foundations of hypothesis testing from a sampling theoretic approach, emphasising the properties of unbiasedness and invariance and methods of constructing tests based on these principles. After considering several asymptotic procedures for testing hypotheses after estimation by Maximum Likelihood, we discuss a class of exact tests and survey the relative merits of these two general approaches.

In chapter 3 we concentrate on two major testing problems in correctly specified linear regression models. Tests against autocorrelation and heteroscedasticity alternatives can each be thought of as seeking sufficient evidence to reject the null hypothesis of a scalar covariance matrix. It is natural, therefore, to consider both problems together. The chapter is divided into three subsections which consider autocorrelation, unconditional heteroscedasticity and conditional hetero-scedasticity respectively. In each case the literature is discussed approximately in order of publication so that an historical perspective is gained.

Chapter 4 abandons the correct specification assumption and surveys the literature concerned with the effects that various mis-specifications have on the properties of tests against autocorrelation alternatives. The chapter is divided into five subsections each of which is devoted to mis-specification of a particular type. We argue that there are important omissions from this literature and we identify several lines of research, some of which are investigated below. Many more interesting research opportunities exist in the literature on testing for heteroscedasticity in mis-specified models but these fall outside the scope of this thesis.

In chapter 5 we take up an important issue concerning the joint modelling of autocorrelation and heteroscedasticity. It is shown in section 5.1 that, although there are two basic modelling options when the heteroscedasticity is not conditional, one of these precludes simultaneous control over the degree of each phenomenon. Eliminating this option, however, still leaves open the precise form of the covariance matrix, two forms of which are possible. Some parallels are drawn with random coefficient models and those with dependent variables which are measured with error. The second subsection of this chapter deals with conditional heteroscedasticity and introduces the two models which have been proposed to accomodate autocorrelation into these variance models. This chapter is essential background for the material of chapters 6 and 8.

The problem of testing for serial independence against the alternative of AR(1) errors, when the errors are unconditionally heteroscedastic, is the subject of chapter 6. This chapter extends work by Giles and Small (1991), and Krämer (1985). After describing the tests to be studied, the technique used to evaluate the exact power functions of each are outlined. Then, using each of the covariance specifications derived in chapter 5, we investigate the boundary regions of the stationary parameter space analytically, establishing several theorems. Using six different sets of data, we proceed by numerically evaluating the power functions of the tests under a variety of forms and degrees of heteroscedasticity. Several valuable conclusions emerge from the research reported in this chapter concerning, in particular, the behaviour of the power functions of the tests considered under very severe degrees of autocorrelation.

The same group of tests is studied in chapter 7, where the true autocorrelation process is assumed to differ from the assumed AR(1) model. The true errors are AR(2), MA(1) and AR(5) in sections 7.2, 7.3 and 7.4 respectively. Exact power functions are used in numerical studies of each form of mis-specification. Analytic results (along the lines of those in chapter 6) are more difficult to establish for AR(2) and MA(1) errors, however, due to the nature of the parameter space. In the AR(2) case most of the stationarity boundary involves two parameters, while MA(1) models have no such boundary, being always stationary. The AR(5) process which is considered in this chapter has several restrictions imposed on the parameters to produce a multiplicative combination of simple first and fourth order AR models, the fifth order term being the interaction between these two processes. In this case each of the boundaries of the stationarity region depends on one parameter only and the analytic part of section 7.4 exploits this aspect. The concluding section of chapter 7 emphasises the major analytic and empirical findings.

The research reported in chapter 8 is of direct relevance to financial econometricians. We consider the problem of testing for serial independence (which amounts to modelling the conditional mean of the dependent variable) in the presence of conditional heteroscedasticity of a particular form due to Bollerslev (1986). This relatively recent variance specification has proved highly successful in capturing well known features of returns on financial assets and has led to an explosion in the numbers of applications published in the empirical finance literature.

Using this literature as a starting point, a set of tests was identified as being of importance. This group includes exact tests, such as those used in the previous chapters, asymptotically justified tests and "robust" tests. This latter group has been formulated specifically with the aim of avoiding the types of problems considered in this thesis. Curiously, there appears to have been no serious attempt to investigate the properties of either the standard (or non-robust) procedures or the newer "robust" tests. The contribution of chapter 8 is to rectify this omission. The nature of the GARCH model is most discouraging from the analytic viewpoint, even without the complication of using asymptotic tests. For this reason test properties are evaluated through a Monte Carlo study using a variety of data. As well as revealing the relative strengths of tests in common use, we consider the implications of sample size and the dependence of our conclusions on the (often unreasonable) assumption of conditionally normal errors.

The conclusions of chapter 8 strongly suggest that the use of exact tests for serial independence is preferable to the currently available asymptotic procedure. Furthermore, the so called robust tests are substantially less powerful than the standard versions on which they are based.

The thesis concludes in chapter 9 with an overview of our most important findings. This is followed by some thoughts on potentially useful further research which might be undertaken on the topics covered here, and on related issues.

CHAPTER 2

HYPOTHESIS TESTING

In this chapter we consider several issues of fundamental importance in the testing of statistical hypotheses. Section 2.1 briefly explores some of the motivations for conducting an hypothesis test in the context of an econometric model and the risks involved in the specification and interpretation phases of modelling. We then attempt to characterise an "optimal" test, in various circumstances, and discuss techniques for restricting the set of tests under consideration. This is covered in section 2.2 and is followed by a discussion of the properties of the maximum likelihood based Wald, Lagrange Multiplier and Likelihood Ratio test procedures. The chapter closes with section 2.4 in which we characterise exact tests and discuss their relative advantages.

2.1 Motivation

Theoretical economic models postulate relationships between several quantities, on the basis of assumed behavioural responses of individuals. By virtue of their multivariate nature, these models are invariably parametric and most formal econometric work therefore includes the estimation of parameters. It is the aim of this section to show that the testing of hypotheses about these parameters is central to the practice of economics, by which is meant the application of economic analysis to real world situations.

We should begin by observing that thorough economic analysis has both theoretical and empirical facets. To paraphrase Zellner (1992), it could be said that measurement without theory is as useless as theory without measurement. The ultimate test of a theory is its ability to predict the future conditional on the information available at present, but what are we to infer from predictive failure? Incorrect prediction may be induced by aberrations of various sorts in the available data, in which case we might wish to defer judgment on the theory. On the other hand, it may be that the theory is simply wrong and should be modified or abandoned. These methodological considerations apply not just to economics, of course, but to science generally. It is the non-experimental nature of economics which sets it apart from most other sciences and severely constrains direct testing of the theory. This poses special problems for empirical workers and strengthens the justification for emphasising the interdependence of theory and measurement in economics.

Hypothesis testing has two roles in the practice of economics, which are often executed simultaneously. These roles concern, first, the specification of the statistical model which is generally based upon economic theory; and second, the use of this model to draw conclusions about economic realtionships. We discuss these roles individually for convenience, while recognising their interdependence in many situations.

In the specification phase of econometric modelling an initial version of the model is established using those data which are thought to be relevant and are available. (On many occasions the intersection of these two categories contains regrettably few variables.) This model is estimated by some appropriate technique and subjected to a series of diagnostic tests which are designed to reveal shortcomings in the initial specification such as non-normal errors, omitted variables or a non-scalar error covariance matrix. Each of the tests used in this phase can reveal information about the statistical adequacy of the model which the careful researcher may use to modify either the form of the model or the method of estimation, or both. There are, of course, many unmentioned opportunities for error in the procedure sketched above, which is intended merely as an illustration of the role of hypothesis testing in the specification of econometric models.

Having arrived at a specification that is apparently acceptable from both the statistical and economic viewpoints, the model is often subjected to another set of hypothesis tests, the results of which are intended to give information either about the relationships between some of the variables used, or about the economic theory on which the model is based.

Some economic theories are indisputably correct (such as negative price elasticities of demand for normal goods) and a contrary empirical finding would cast doubt on the validity of the estimate. In other cases, such as the characterisation of the determinants of business fixed investment, the disputes between conflicting theories can only be settled empirically. This must be done by reducing the theory to a single testable hypothesis and examining the data for evidence that is sufficient to justify rejection of the hypothesis. Ideally, one theory's model would be nested within the other so that the choice of theory could be made of the basis of the outcome of a test of the validity of a set of restrictions. The enduring nature of some theoretical disputes is, however, indicative of the fact that such problems do not always yield to this approach. In cases where the nesting technique is not feasible the competing models can be interrogated individually, again using hypothesis tests. Alternatively, a more formal model-selection procedure might be invoked, such as those suggested by Akaike (1973), Cox (1962) or Pesaran (1974).

Having established that the testing of hypotheses is a crucial element of practical economics, we now consider some relevant aspects of the application of hypothesis tests. A variety of procedures are available for approaching most testing problems, so the selection of the appropriate test must precede any application. The next section introduces some optimality criteria which provide some guidance for test selection. Other issues, such as the choice of suitable significance levels and the robustness of test procedures, are discussed for selected cases in later sections of this thesis.

2.2 Foundations

In this section the statistical theory which underlies classical hypothesis testing is formalised, with the aim of characterising an "optimal" test. The exposition draws on the classic textbook by Lehmann (1986) although the coverage is necessarily much more limited. In the interests of clarity the theory is discussed with only occasional reference to the measure theoretic concepts on which classical hypothesis testing is based. Discussion of inference from a Bayesian viewpoint is a major topic which is not treated here as it falls outside the scope of this thesis.

The starting point for our discussion is the sample space A which is an abstract representation of all possible outcomes. For our purposes it will be convenient to think of an outcome as a particular realisation of the set of economic variables under study. The sample space is then the set of all possible such realisations. We wish to be able to make statements about the (possibly conditional) probability of particular groups of outcomes, which are necessarily subspaces of A. To enable such statements to be made it is necessary that the concept of probability is defined over A and it's subsets, S_i. We must therefore require that $P(0)=0$, $P(A)=1$ and that the joint probability, $P(A)$ is equal to the sum of the individual probabilities, $\sum_{i=1}^{n} P(S_i)$, $i=1,2,\ldots n$. If these conditions are met the sets A and S_i are said to be measurable[1].

Suppose that the problem of interest concerns the true value of a parameter (vector) θ, and is stated as a choice between the null hypothesis H_0: $\theta \in \Omega_0$ and the alternative hypothesis H_1: $\theta \in \Omega_1$. We wish to base a decision between these hypotheses on the observed value of the random variable X, the distribution of which is assumed to belong to the class $F=\{F_\theta, \theta \in \Omega\}$. The null and alternative hypotheses are assumed to be mutually exclusive, as are the corresponding subsets of Ω which are denoted Ω_0 and Ω_1 respectively. The test statistic which is employed to decide between H_0 and H_1 is a single realisation, x, of X and associated with it is some probability $\phi(x)$ of rejecting H_0. In what follows we shall refer to ϕ as a test since, given x, ϕ is a function which determines the test outcome.

It is clearly desirable to construct a test which maximises the probability of making the correct decision. The two possible decisions (reject or do not reject H_0) are mutually exclusive, however, and each incurs an error under some circumstances. These errors, therefore, cannot be simultaneously controlled. The standard classical solution to this dilemma is to select some value α to form an upper bound on the probability of incorrectly

[1] Strictly speaking the sets are still measurable without $P(A)=1$; this condition restricts the possible measures to probability measures.

rejecting H_0. This value is known as the level of significance, or size (of the region consistent with H_0) of the test. We require that $\phi(x) \leq \alpha$ when $\theta \in \Omega_0$, or equivalently:

$$(1) \qquad P_\theta(X) \leq \alpha \quad , \quad \text{for all } \theta \in \Omega_0 \ .$$

This condition forms a constraint on the testing problem which, given (1) amounts to finding a ϕ for which the probability of erroneously failing to reject H_0 is minimised. This can be expressed as:

$$(2) \qquad \max_{\phi} \ \phi(x) \quad \text{for all } \theta \in \Omega_1, \text{ subject to (1)} \ .$$

In general, further structure is needed before this problem can be solved. This is because Ω_1 usually contains many distributions so that the solution to (2) depends on the true value of θ, which is of course unknown. There are several ways of obtaining extra structure for (2) but a discussion of these should be preceded by considering the special case in which both H_0 and H_1 are simple (*i.e.*, there is only one distribution in each of Ω_i, i=0,1). In this case the following result, known as the fundamental lemma of Neyman and Pearson (1933), establishes the existence of a solution to (1) and (2) and provides necessary and sufficient conditions for that solution.

Theorem 2.2.1

Let p_0 and p_1 be the densities of two probability distributions. For testing $H_0:p_0$ against $H_1:p_1$ there exists a test ϕ and a constant k such that

$$(3) \qquad E_0\left\{ \phi(X)=\alpha \right\} \text{ and } \phi(x)=\begin{cases} 1 \text{ iff } p_1(x) > kp_0(x) \\ 0 \text{ iff } p_1(x) \le kp_0(x) \end{cases}.$$

The test which satisfies (3) for some k is the most powerful size α test of p_0 against p_1. The converse also applies unless there exists a test of size $\beta < \alpha$ with power equal to unity.

Proof: See Lehmann (1986) pp.74-76.

From Theorem 2.2.1 a most powerful test always exists when both hypotheses are simple. In practical applications, however, this condition is rarely met, since testable hypotheses generally involve at least one composite hypothesis, so that Ω_i contains more than one distribution for some i. If H_1 is composite then a single size α test which solves (1) and (2) for each and every $\theta \in \Omega_1$ is said to be Uniformly Most Powerful (UMP). Composite alternatives can be either one-sided, as in H_1: $\theta > \theta_0$, or two-sided (H_1: $\theta \neq \theta_0$). The choice of alternative hypothesis is at the discretion of the researcher and depends on various aspects of the application.

Without further restrictions on the class of tests, relatively few problems admit a UMP test. One notable exception occurs when testing a single real valued parameter for which the family of densities $p_\theta(x)$ has a monotone likelihood ratio, $p_{\theta_1}(x)/p_{\theta_0}(x)$. Provided that the cross partial derivative $D_2=\partial^2\log(p_\theta(x))/\partial\theta\partial x$ exists, a necessary and sufficient condition for $p_\theta(x)$ to have a monotone likelihood ratio in x is that $D_2\geq 0$ for all θ and x.

Despite the relatively rare occurrence of true UMP tests there are many cases in which the non-existence of such a test does not unduly restrict the construction of "ideal" tests. The key to further progress is to impose reasonable conditions on such a test and then restrict attention to those tests which meet the conditions. Perhaps the most reasonable such condition is the requirement that a test be "unbiased". An unbiased test is one for which the rejection probability for any element of the class of alternative hypotheses is not less than the size of the test $(\phi(x)|\theta\in\Omega_0)$. In the case of a biased test, for some $\theta \in \Omega_1$ one is less likely to reject H_0 when it is false than when it is true. This is clearly an undesirable outcome and justifies the unbiasedness restriction.

Although it is obviously sensible to consider only unbiased tests, this restriction as it stands is not particularly helpful in the construction of "optimal" unbiased tests. It is easily shown, however, that unbiasedness implies that $E_\theta\left\{ \phi(x) \right\} = \alpha$ for all θ on the boundary between Ω_0 and Ω_1, provided that the power function of the test is continuous in θ. For a given testing problem all such tests are said to be "similar" (on the boundary of Ω_0 and Ω_1). In this form the unbiasedness restriction can provide a valuable reduction in the class of tests worthy of consideration.

This reduction is conveniently revealed through consideration of testing problems involving some nuisance parameter, n, which affects the distribution of X. Notice that on the Ω_0, Ω_1 boundary the rejection probability of all similar tests, conditional on n, is equal to α and independent of θ. The problem is therefore reduced to one of testing a simple hypothesis for each n.

The principle of unbiasedness is not always useful in the construction of optimal tests, however. In some cases, such as when the null and alternative hypotheses do not adjoin, this is because the concept of similarity on the boundary is meaningless.[2] In other problems, clearly defined boundary regions are insufficient. There seems, for example, to be no UMP unbiased (UMPU) test of $H_0: (\mu/\sigma) \leq \theta_0$ against $H_1: (\mu/\sigma) > \theta_0$ in a $N(\mu, \sigma^2)$ distribution. In cases such as these, recourse to a further principle, that of "invariance", can sometimes allow further progress.

When the testing problem is unchanged by a group (in the formal sense) of transformations it is natural to require that the outcome of any hypothesis test be similarly unaffected by transformations within the group. To use a familiar example, the

---

[2] An example is the problem of choosing between two non-nested models; see Pereira (1977) for a survey of this literature.

OLS estimator of the parameter vector in a classical linear regression model is invariant to the ordering of the data points. Any reasonable test of the true value of a subset of the parameters should not, by the invariance principle, produce outcomes which are dependent on the data order.

To formalise this, suppose that $X \sim P_\theta$, $\theta \in \Omega$ and g is a 1:1 transformation of the sample space $\mathcal{X}$ onto itself. Let $gX=gx$ when $X=x$, with $gX \sim P_{\theta'}$, $\theta' = \bar{g}\theta \in \Omega$. We say that Ω is invariant under g if

(4) $$\bar{g}\Omega = \Omega \quad .$$

Equation (4) implies that $\bar{g}\theta \in \Omega$ for all $\theta \in \Omega$ and that, for any $\theta' \in \Omega$, there exists some $\theta \in \Omega$ such that $\bar{g}\theta = \theta'$. It does not, however, imply that the testing problem itself is invariant under g since it leaves open the possibility that for some $\theta \in \Omega_0$, $\bar{g}\theta \in \Omega_1$. To rule this out we say that the testing problem is invariant to g iff (4) holds and

(5) $$\bar{g}\Omega_0 = \Omega_0 \quad .$$

For problems which satisfy (4) and (5) the principle of invariance suggests that we should seek a test which is optimal among the class of tests for which

(6) $$\phi(gx) = \phi(x) \ , \qquad \text{for all } x \quad .$$

This, however, is of limited practical value for the construction of an invariant test, in the same way that the principle of unbiasedness does not directly guide the construction on an unbiased test. We discussed above the role of similarity on the boundary in overcoming the latter problem, For the same reasons, we now introduce the concept of a maximal invariant. If two points x_1, x_2 satisfy $gx_1=x_2$ for some g within the group G, then the orbits[3] of G are partitions of the sample space. For a function M to be invariant under G it follows that $M(gx) = M(x)$ must be constant for all $g \in G$. If, in addition, each orbit takes on a different value then M is said to be a maximal invariant[4] with respect to G. The utility of maximal invariants lies in the direct link between a maximal invariant and an invariant test, which is provided by the following theorem.

Theorem 2.2.2

If M is a maximal invariant with respect to G then a necessary and sufficient condition for a function f to be invariant (w.r.t. G) is that there exists a function h for which $f(x) = h(M(x))$ for all x.

Proof:

If f is invariant and $M(x_1) = M(x_2)$, then $gx_1 = x_2$ for some g, so that $f(x_2) = f(x_1) = h(M(x_2))$ for some h. To see that the theorem also provides a sufficient condition, notice that if $f(x)=h(M(x))$

---

[3] A single orbit of G is the locus of points gx_i formed by applying each $g \in G$ to x_i.

[4] Clearly, all maximal invariants are constant over the same sets.

for all x, then $f(gx) = h(M(gx)) = h(M(x)) = f(x)$. #

Theorem 2.2.2 shows that all functions which are invariant under G are themselves functions of a maximal invariant M. This statement can be extended to invariant tests by imposing the additional requirement that the class $\mathcal{B}$ of measurable sets B satisfies $M^{-1}(B) \in \mathcal{A}$, the sample space.

To demonstrate the use of the the invariance principle, consider the sample $X_1, \ldots, X_n$ from the $N(\mu, \sigma^2)$ distribution and the hypotheses H_0: $\sigma \geq \sigma_0$ and H_1: $\sigma < \sigma_0$. This problem is invariant to transformations of the form $X_i' = X_i + c$, $-\infty < c < \infty$, and a maximal invariant is $S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2$. All invariant tests are therefore functions of S^2. Furthermore, since S^2 has a monotone likelihood ratio in x, it also provides the basis for a UMP Invariant (UMPI) test of H_0.

The relative merits of the principles of unbiasedness and invariance depend on the testing problem under study. For some problems, such as H_0: $(\mu/\sigma) \leq \theta_0$ in a $N(\mu, \sigma^2)$ distribution, invariance is useful while unbiasedness is not. Unbiasedness does, however, have one strong unconditional advantage which should be mentioned. If a UMPU size α test exists, then no other size α test exists for which the rejection probability is at least as great for all $\theta \in \Omega_1$ and greater for one such θ. Any UMPU test is therefore admissible. This is not necessarily true of a UMPI test, although it may be satisfied in particular applications.

We have shown how the principles of unbiasedness and invariance can be useful, first in reducing the class of tests from which an optimal member is chosen, and second in suggesting a methodology for making this choice. In the new research reported later in this thesis the optimality properties of the tests investigated, conditional on a correct model specification, will be reported with reference to the foregoing discussion. We now discuss a class of tests which are derived from the properties of a particular parameter estimator, and draw parallels between these tests and the topics covered above.

2.3 LR, LM and Wald Tests

In this section we consider three very general methods of test construction, each of which is based upon the maximum likelihood estimates of a regression model. The relation of these tests to the maximum likelihood estimator is useful in revealing their links to each other, and also indicates that their use is confined to models which yield to this estimation technique. This rules out the use of these tests in models estimated by other methods, perhaps due to the excessive computational burden imposed by maximum likelihood estimation[5]. Newey and West (1987) provide analogues for these tests under efficient generalised

[5] Some rational expectations models fall into this category owing to the complicated serial correlation and heteroscedasticity induced in the disturbances by the nature of the model.

method of moments estimation which provides an attractive alternative to maximum likelihood estimation in such cases. This topic, however, falls outside the scope of this thesis and will not be considered further. In general, the tests considered in this section have (large sample) asymptotic justification but, in certain circumstances, can correspond to optimal exact tests. The distinction between exact and asymptotic tests will be clarified in the final section of this chapter.

To introduce the material of this section we consider the implications of an infinitely large sample size. This would be equivalent to having full knowledge of each variable and should completely eliminate uncertainty about the relationship under study. Under such circumstances we would hope that the power of an hypothesis test would be unity. A test which has this property satisfies

$$\phi_T(x) = \lim_{T \to \infty} \phi(x) = 1 \; ; \qquad \text{for all } \theta \in \Omega_1$$

and is said to be consistent. This is a very weak property however, and consequently it does not appreciably narrow the class of tests worthy of consideration. For asymptotic tests the usual criterion for test selection is the ability to correctly discriminate between θ_0 and values "close" to θ_0. By suitably restricting the range of Ω_1 we can ensure that the relevant test statistics are $O_p(1)$ rather than $O_p(n)$ so that the asymptotic power is less than unity. In fact, an infinite sample size drives such alternatives to the null hypothesis and for this reason they are known as local alternatives.

A major advantage of this approach can be seen by considering the non-linear hypothesis $H_0:g(\theta)=0$, where g is a vector of functions. Let θ_0 be the true value of θ if H_0 is true, and let θ' be some local alternative lying between θ and θ_0. Consider the following first order Taylor series approximation of $g(\theta)$ about θ_0 (assuming the existence of $G(\theta)=\partial g/\partial\theta$):

$$g(\theta) = g(\theta_0)+G(\theta')(\theta-\theta_0).$$

In the neighbourhood of the null θ is close to θ_0 so that $G(\theta')$ is close to $G(\theta_0)$. Furthermore, under the null $g(\theta_0) = 0$ so that $g(\theta)$ reduces to $G(\theta_0)\theta_0$. Thus we see that for local alternatives every non-linear hypothesis can be approximated by a linear form.

To discuss the estimation of θ we begin by specifying that the observations $y_0,y_1,\ldots,y_n$ (which may be vectors) have density functions of the form $f_t(y_t|Y_{t-1};\theta)$; where $Y_{t-1}=(y_{t-1},y_{t-2},\ldots,y_0)$ is the known history of y, the form of $f_t(.)$ is known, and θ is a k dimensional parameter vector (or matrix as necessary). Conditional on the data, the log-likelihood function of θ is

$$\ell(\theta) = ln(L(\theta)) = \sum_{t=1}^{n} \ln(f_t(y_t|Y_{t-1};\theta) \tag{7}$$

In light of the discussion on local alternatives we assume that the hypothesis of interest can be written as a set of s exact linear restrictions on the parameter vector:

$$H_0:\ R\theta - r_0 = 0 \tag{8}$$

where R is an sxk matrix of known constants and the elements of the vector r_0 are also known. This specification is sufficiently general to accomodate many of the testing problems encountered in econometrics. Assume also that the parameter space Ω is closed, bounded and of finite dimension, and that the true parameter vector θ_0 is interior to Ω.

The score vector $D(\theta) = \partial l(\ell)/\partial\theta$ is assumed to be continuous the neighbourhood of θ_0, as is the Hessian matrix $D^2(\theta)=\partial^2\ell(\theta)/\partial\theta\partial\theta'$. The negative of the expectation of $D^2(\theta)$ is known as the (Fisher) information matrix $I(\theta)$ and is assumed to be positive definite in the neighbourhood of θ_0. To ensure the consistency of maximum likelihood estimators we assume that the eigenvalues of $I(\theta_0)$ approach infinity as the sample becomes asymptotic, thus increasing the sample information without limit. In the neighbourhood of θ_0 we require that plim $-I^{-1}(\theta)D^2(\theta)$ is the kxk identity matrix I_k. Finally some regularity conditions (see Crowder (1976)) on the score vector are needed to allow the application of a martingale central limit theorem to the asymptotic distribution of $D(\theta)$[6].

The restricted maximum likelihood estimator (MLE) of θ is the parameter vector which maximises (7) subject to (8) and will be denoted $\hat{\theta}$, while the corresponding unrestricted estimator is $\tilde{\theta}$. Under our assumptions $\tilde{\theta}$ converges asymptotically to the $N(\theta_0, I^{-1}/T)$ distribution, (and hence is consistent) irrespective

[6] In essence we require the asymptotic equivalence of a conditional (on Y_{t-1}), and an unconditional arbitrarily weighted sum of squared scores.

of the validity of the restrictions. It is therefore clear that

$$T(\tilde{\theta} - \theta_0)' \; I(\theta_0) \; (\tilde{\theta} - \theta_0) \;\; \to \chi^2_s \; , \quad \text{if } H_0 \text{ is true.}$$

(Here $\to$ means "converges in distribution to") Furthermore, $I(\theta_0)$ can be replaced by any consistent estimator, such as that based on the MLE, and since $R\theta = r_0$ under H_0 we see that:

$$t_W = (R\tilde{\theta} - r_0)'(R \; I^{-1}(\tilde{\theta})R')^{-1}(R\tilde{\theta} - r_0) \;\; \to \;\; \chi^2_s \quad \text{if } H_0 \text{ is true.}$$

This is known as the Wald statistic and, while applicable to a wide range of testing problems, it is particularly suited to those for which estimation is significantly easier for the unrestricted (or alternative) model than for the restricted (or null) model.

The Wald test is easily shown to be an asymptotic approximation to the familiar t and F tests in the context of a classical linear regression model. It can also be implemented in non-linear regressions and with non-linear constraints, however in the latter case the test is not invariant to equivalent forms of the constraints (see Gregory and Veall (1985), Lafontaine and White (1986) and Phillips and Park (1988)).

It was mentioned in the previous section that when the testing problem concerns only one real valued parameter with a likelihood ratio which is monotonic in x, then a UMP test is provided by $\phi(x)$. The natural extension of this to a vector of parameters is embodied in the following Likelihood Ratio Test which is based on the difference between the maximised likelihood functions under the null and the alternative hypotheses. Consider the following Taylor series expansion of $\ell(\hat{\theta})$ about $\tilde{\theta}$:

$$\ell(\hat{\theta}) = \ell(\tilde{\theta}) + (\tilde{\theta}-\hat{\theta})'D\ell(\tilde{\theta}) + \tfrac{1}{2}(\tilde{\theta}-\hat{\theta})'D^2\ell(\tilde{\theta})(\tilde{\theta}-\hat{\theta}).$$

Noting that $D\ell(\tilde{\theta})=0$ and defining λ such that $\ln\lambda=\ell(\hat{\theta})-\ell(\tilde{\theta})$ we can rewrite this expansion as:

$$t_{LR} = -2\ln\lambda = (\tilde{\theta}-\hat{\theta})'(-D^2\ell(\tilde{\theta}))(\tilde{\theta}-\hat{\theta}).$$

This statistic has an asymptotic χ^2_s distribution under the null (see Theil (1971, p.396) for a proof) implying rejection when $\ell(\tilde{\theta})$ is sufficiently large compared with $\ell(\hat{\theta})$. In the classical linear regression model the LR test of a single restriction is equivalent to the t test and in this context the one-sided version of the test is UMP, providing a direct link back to the Neyman-Pearson Lemma.

A third approach to hypothesis testing is based on the restricted model only. Estimation of θ subject to (8) is achieved by the maximisation of the Lagrangean function:

$$L^* = \ell(\theta) + \lambda'(R\theta - r_0)$$

where λ is an sx1 vector of Lagrange multipliers associated with the parameter constraint. The Lagrange Multiplier (LM) test is based on the statistic

$$t_{LM} = \hat{\lambda}' \; R \; I^{-1}(\hat{\theta}) \; R' \; \hat{\lambda}$$

where a ^ indicates values which maximise L*. An equivalent representation, $t_{LM} = D\ell(\hat{\theta})'I^{-1}(\hat{\theta}) \; D\ell(\hat{\theta})$, was denoted the score test statistic by Rao(1973). Each representation is asymptotically distributed as a χ^2_s variate under the null hypothesis.

It is clear that for any testing problem t_W, t_{LR} and t_{LM} have the same asymptotic distribution. The method of Silvey

(1959) can be used to show that they also share the same distribution for tests against a sequence of local alternatives. These three tests are therefore said to be asymptotically equivalent. It can also be shown (see Engle (1984)) that, under the assumptions made above with respect to the likelihood function, each of the test statistics is asymptotically a function of a maximal invariant and is therefore asymptotically locally most powerful invariant (LMPI). As there are no differences between the tests which can be found using asymptotic theory, the choice criteria is generally that of computational convenience, although finite sample considerations are also important.

Because the LM test only requires estimation of the null model it is particularly useful for applications in which the unrestricted model is difficult to estimate for any technical or informational reason. In many cases the null model is most appropriately estimated by a Generalised Least Squares (GLS) estimator due to the block diagonal structure of the information matrix. For such models an asymptotically equivalent representation of t_{LM} is given by TR_0^2 where R_0^2 is the uncentered R^2 from a regression of the residuals from the null model estimation, $\tilde{u}$, on the derivatives $\partial\tilde{u}/\partial\theta$ evaluated at the restricted estimate, $\hat{\theta}$.

If we continue to confine attention to problems which are able to be addressed in a GLS framework then some further connections between these three procedures emerge. First it can be shown that the distinction between the tests hinges on the choice of estimator for the error covariance matrix, the Wald test using the unrestricted form, the LM test using the restricted form and the LR test using both. It is therefore to be expected that, although asymptotically equivalent, some finite sample differences might emerge between these tests. Building on the work of Savin (1976), Breusch (1979) established that when testing linear restrictions in linear models the following inequality exists between the test statistics:

$$t_W \geq t_{LR} \geq t_{LM}.$$

The second part of this ordering also holds for non-linear restrictions and non-linear models. This relationship, however, holds under both H_0 and H_1 and therefore says nothing about the relative power of the tests. Indeed, evidence from Evans and Savin (1982) suggests that the size corrected powers are approximately equal across the tests.

2.4 Exact Tests

The discussion in the previous section concerned several very general methods of test construction, at least one of which is applicable in virtually any econometric context. It was made clear that the statistics discussed converge to χ^2 distributions only as the sample size approaches infinity. For some testing problems a superior alternative exists in the form of a optimal (in some sense) statistic for which the finite sample distribution is known. Provided that the significance points of this distribution are computable, such a test is usable and is known as an exact test.

Familiar examples of exact tests are those t and F distributed ratios associated with tests of restrictions on the coefficients in the classical linear regression model. For these tests, the null distribution of each statistic (conditional on the validity of the assumptions underlying the model) is tabulated for given degrees of freedom and significance level. Exact tests against non-scalar covariance matrices include the exact version of the Durbin-Watson test as well as the class of point optimal tests for autocorrelation (see chapter 3 below) and heteroscedasticity (Evans and King (1985b)).

The major advantage of an exact test *per se* is the elimination of the uncertainty surrounding the distribution of the test statistic in a particular application. There are, however, two general conditions under which this is not necessarily an advantage. The first and most important is the almost inevitable mis-specification of the regression model along with the consequent violation of assumptions which underly exact tests. In this case the assumed "exact distribution" may be very different from the actual distribution of the test statistic, leading to possibly incorrect inferences drawn from tests with poor power properties. This issue is taken up in greater detail in chapter 4, which considers the problem of testing for autocorrelation in mis-specified regressions. The second condition under which exact tests are not necessarily advantageous is when the sample is "sufficiently large" for asymptotic test statistics to converge to their limiting distributions. The definition of what constitutes a sufficiently large sample can only be determined empirically and is most conveniently tackled through the estimation of response surfaces (see Hendry (1984)) for individual tests under particular data conditions.

This chapter has discussed material which is basic to the testing of hypotheses in econometrics. We now turn to a more detailed discussion of two particular testing problems.

CHAPTER 3

TESTING REGRESSION ERRORS FOR THE PRESENCE OF A SCALAR COVARIANCE MATRIX

In this chapter the literature on two major econometric testing problems is surveyed. The discussion is based on the following linear regression model:

$$(1) \qquad y = X\beta + u$$

where X is a matrix of n observations on k linearly independent non-stochastic regressors, β is a (k×1) vector of parameters to be estimated and u is an nx1 vector of random disturbances. We assume that $u \sim N(0,\sigma^2\Omega)$, where Ω is non-random, positive definite and symmetric, and we wish to test H_0: $\Omega = I_n$ against various non-scalar alternative forms of Ω.

It must be emphasised that throughout this chapter, the specification of (1) is assumed to be correct in the sense that all relevant regressors are included in X, the parameters in β are constant throughout the sample, the true functional form is linear, and the underlying probability distribution is normal. Furthermore, in discussing tests of Ω we shall in general confine ourselves to alternatives which are assumed to be correctly formulated. This final assumption will be relaxed in specific ways beginning in the next chapter.

The covariance matrix Ω contains information about the properties of u which can be broadly bisected into those related to autocorrelation and heteroscedasticity. These phenomena are discussed separately although there are strong links between them which are particularly apparent in the recent parameterisations of heteroscedasticity, and will be covered below. The chapter begins with a survey of tests for autocorrelation in section 3.1. Tests for heteroscedasticity are discussed in section 3.2 and will be divided into those concerned with unconditional heteroscedasticity, which are discussed in section 3.2.1 while conditional heteroscedasticity tests are covered in section 3.2.2. Some concluding comments are made in section 3.3.

3.1 Tests of Serial Independence

In this section we discuss the literature on testing for serial independence in a correctly specified linear regression model. The coverage is restricted to "simple" tests (as distinct from joint tests designed to detect autocorrelation in conjunction with some other departure from the standard assumptions listed in connection with (1) and known as the "classical" assumptions) and includes neither linear simultaneous equations models, nor pre-testing issues. These restrictions on our analysis are imposed for two reasons. First, our major aim in this thesis is to establish some of the consequences of testing for serial independence while ignoring the possibility that other classical assumptions are violated. We are therefore primarily interested in simple tests. Second, hypothesis testing in simultaneous equations models and the consequences of pre-testing for serial independence require knowledge of the material in this section in any case, and each is a major extension in its own right. For a discussion of the econometric issues involved in the former see Godfrey (1978) or Harvey and Phillips (1980); an up to date survey of the state of pre-testing knowledge is provided by Giles and Giles (1993). The material of this section is discussed in approximately the order of publication, hence we begin with the AR(1) process and progress to higher order models.

Most of the early research on testing for autocorrelation in an econometric model considered the problem of testing $\rho=0$ in the following AR(1) model:

$$(2) \qquad u_t = \rho u_{t-1} + \varepsilon_t \ ; \qquad \varepsilon_t \sim IN(0,\sigma^2), \quad t=1,\dots,n.$$

where u_t is the disturbance in (1) at time t. Initial work by Koopmans (1942), however, considered the problem of detecting an AR(1) process in an observed series, rather than the unobserved random disturbances of (1).

Explicitly considering serially correlated errors in (1), Anderson (1948) showed that for some X matrices[1], when testing

[1] Those (possibly random) matrices for which the column space of X is spanned by eigenvectors of Θ.

$H_0:\lambda=0$ against $H_1:\lambda>0$ assuming that $u \sim N(0,\sigma^2(I+\lambda\Theta)^{-1})$[2], a UMP similar test is provided by rejecting H_0 for small values of $r = \hat{u}'\Theta\hat{u}/\hat{u}'\hat{u}$. Here $\hat{u}$ is the OLS residual vector and it is assumed that Θ is a known matrix, σ^2 and λ are unknown scalars and the covariance matrix of u is positive definite. Anderson also showed that no UMP similar test of independence against a positive alternative exists for the special case of stationary AR(1) errors in (1).

Despite this finding the detection of AR(1) errors became the subject of much research interest. Following closely behind the influential papers of Cochrane and Orcutt (1949) and Orcutt and Cochrane (1949) came work by Durbin and Watson (1950, 1951) which provided the basis for what is still the most commonly used test for serial independence against the alternative of AR(1) errors. Durbin and Watson (henceforth DW), studied the distribution of $d = \hat{u}'A\hat{u}/\hat{u}'\hat{u} = u'MAMu/u'Mu$ where $M=I-X(X'X)^{-1}X'$ and A is any real symmetric matrix. The following choice of A drew on Anderson's (1948) work:

$$A = \begin{bmatrix} 1 & -1 & 0 & . & . & . & . & 0 \\ -1 & 2 & -1 & & & & & . \\ 0 & -1 & 2 & & & & & . \\ . & & & . & & & & . \\ . & & & & . & & & . \\ . & & & & & . & & 0 \\ . & & & & & & 2 & -1 \\ 0 & . & . & & & 0 & -1 & 1 \end{bmatrix}$$

[2] This form of covariance matrix (except for the North-West and South-East elements) can incorporate any AR(p) process under weak conditions on the sample size (see Faust (1992)).

This differencing matrix is equivalent to the inverse of the covariance matrix of the u_t of (2) when $\rho=1$. Using this matrix, DW demonstrated that their test is approximately UMP similar when the columns of X are linear combinations of the eigenvectors of A.[3] Finding no expression for the density of r, DW were nevertheless able to show how bounds could be placed upon the exact critical value of the test, for a given nominal size and conditional on the dimensions of the X matrix. This was achieved through the establishment of a lemma using the eigenvalues of A, but is now of only theoretical interest due to the availability of computer algorithms capable of locating exact cdf probabilities with high levels of accuracy. This topic will be discussed further in chapters 4 and 6.

During the two decades following the original DW papers several authors devised approximations to the true DW critical value. This work has been outdated by the readily available exact version of the DW test and will not be treated here. In a modification of the DW test, King (1981) replaced the top left and bottom right elements of A with 2's and found the resulting test to be a true Locally Best Invariant (LBI)[4] test at the origin (*i.e.* under the null hypothesis) against one-sided

[3]The approximation is due to a minor modification made to the joint density function of a stationary AR(1) process.

[4]A LBI test has the most steeply sloped power function within the class of invariant tests, in the neighbourhood of a particular point in the sample space.

alternatives in either direction.

The DW test is not valid for regressions fitted through the origin although operational procedures for such models are provided by DW (1951) and Krämer (1971). The test is similarly invalid when one or more lags of the dependent variable are present. For these cases several alternatives are available including Durbin's h test (Durbin (1970)), the asymptotically equivalent LM test (see Breusch (1978)) and Wald test (Dhrymes (1971)), and a point optimal test due to Inder (1985).

The tests discussed so far are all based on the OLS residual vector, $\hat{u}$. However, $\hat{u} = Mu$, so that under the classical assumptions the residuals are generally serially correlated and heteroscedastic. Furthermore, use of $\hat{u}$ in testing for the serial independence of u, may lead to over-rejection or under-rejection of the null hypothesis. Theil (1965) initiated a further group of tests which avoid this problem by defining a linear unbiased residual vector with a scalar covariance matrix (or LUS residual vector), u^*, by means of the transformation $u^* = By$. Here B is any non-stochastic nxm matrix such that $B'X=0$ and $B'B=I_m$ and $m \leq (n-k)$. To select the most suitable LUS vector from the infinite number of choices Theil defined the best LUS (or BLUS) vector to be the one which minimises the expected sum of squared errors for a chosen n-k observations. This, however, still leaves the problem of which n-k observations to select. For tests of serial independence Theil observed that the BLUS residuals used should be consecutive and recommended dropping k disturbances from either the beginning or the end of the sample, with the minimisation rule being used to choose between these two options.

Another approach to the problem of serial correlation of regression residuals uses the standardised prediction errors from "recursive" estimation of (1). This procedure begins with OLS estimation of β using only the first k observations. The general form of this estimator is

$$b_t = (X_t'X_t)^{-1}X_t'y_t^* , \quad t=k,\ldots,T$$

where $X_t = (x_1,x_2,\ldots,x_t)'$ and $y_t^* = (y_1,y_2,\ldots,y_t)'$. The corresponding estimator based on t+1 observations may be obtained[5] from b_t as:

$$b_{t+1} = b_t + (X_t'X_t)^{-1}x_{t+1}\tilde{v}_{t+1}/f_{t+1}$$

where $\tilde{v}_{t+1} = (y_{t+1} - x_{t+1}'b_t)$ is the one step ahead prediction error and $f_{t+1} = 1+x_{t+1}'(X_t'X_t)^{-1}x_{t+1}'$

The vector of recursive residuals is defined as $v_t = \tilde{v}_t/\sqrt{f_t}$ and was shown to be a LUS vector by Phillips and Harvey (1974) who found little power difference between tests based on BLUS and recursive residuals. Other forms of LUS residuals include vectors constructed using Householder matrices which are known as LUSH residuals (see Golub (1965)) and vectors augmented with the OLS residuals (Tiao and Guttman (1967)).

The available evidence on the power of LUS based tests for serial independence suggests that they are generally inferior to the exact DW test. Koerts and Abrahamse (1969) found that Theil's

[5] See Harvey (1981) for a proof.

BLUS test has lower power than the exact DW test, a conclusion supported by L'Esperance and Taylor (1975). Several authors have found that BLUS based tests are also less powerful than those using LUSH residuals (see Fraser, Guttman and Styan (1976) or Ward (1973)). Because of the generally inferior performance of tests based on transformed residuals, and their very infrequent use in the applied econometrics literature, such tests are not included in the new research reported in this thesis.

An alternative approach to the same problem uses residuals which are Linear, Unbiased and with a Fixed covariance matrix, or LUF. This work is surveyed by Dubbelman (1978), while King (1979) shows that every LUF based test has an exactly equivalent LUS counterpart. While admitting the possibility that a practical and superior alternative to the DW test may exist within the class of LUS based tests, such a test is as yet unknown (except insofar as the Kadiyala-based tests discussed below fit into this category) and these tests will therefore not be considered further.

In the same category is a group of non-parametric tests which focus on patterns in consecutive residuals such as turning points (Kendall (1976)), runs of positives or negatives, or sign changes (Geary (1970)). These tests have unreliable sizes and poor power relative to the DW test and were originally suggested for their computational ease and conclusivity, neither of which is an advantage over the modern DW test.

A more promising line of research has produced several tests against AR(1) errors which are more powerful than the exact DW test for particular values of ρ in (2). Kadiyala (1970) considered testing the H_0: $\Omega = I_n$ against the alternative of a known, non-scalar, positive definite matrix. Noting Lehmann and Stein's (1948) finding that, for an observed series, the most powerful test of this general hypothesis is provided by rejecting for small values of $\omega_0 = u'\Omega u/u'u$, Kadiyala sought a suitable observable series to replace the unobservable u. His choice was essentially a LUS vector, $v = P_1\hat{u}$, where P_1 is an (n-k)xk matrix with rows[6] comprising the eigenvectors corresponding to the unit eigenvalues of M. Clearly v is multivariate normal with zero mean and $E(vv')=\sigma^2 P_1\Omega P_1'$ under the alternative, or $\sigma^2 I_n$ under the null. The most powerful test of H_0 against H_a:$\Omega=I_n$ is thus given by rejection for small values of $\omega_0 = v'(P_1\Omega P_1')^{-1}v/v'v$.

A useful relation was established by King (1980), who showed that $\omega_0 = \tilde{u}'\Omega^{-1}\tilde{u}/\hat{u}'\hat{u}$ where $\tilde{u}$ is the GLS residual vector assuming $E(uu')=\Omega$. Using this result it is easily shown that, for the problem of detecting AR(1) errors with a specific value of ρ, Kadiyala's test corresponds to Berenblut and Webb's (1973) test, the Likelihood Ratio Observable (LRO) test of Fraser, Guttman and Styan (1976) and to the Point Optimal test of King (1985). These tests are all LBI tests in particular regions of the ρ space and will be discussed in greater detail in the subsequent chapters of this thesis.

[6] Strictly the form of P_1 is more general, being restricted to matrices with rows which are orthogonal to each other and to the columns of X.

Before considering higher-order AR processes we observe that it is possible to construct LR, Wald and LM tests of serial independence against an AR(1) alternative. There have been several papers which have considered this approach (see, for example Dent (1973) or Schmidt and Guilkey (1975)) but found no advantage to be gained. As suggested in the previous chapter, there is no reason to use an asymptotic test if an exact test with suitable power properties exists.

The Kadiyala (1970) based locally MPI tests mentioned above are successful, in part, because of the highly restrictive nature of the alternative hypothesis. In AR(1) applications Ω is a function of one parameter (ρ) only and the structure of Ω, conditional on ρ, is well known. To date, however, no generalisation to arbitrarily higher order processes is available. In certain special cases, such as testing against simple AR(p) processes[7], exact tests of independence against such alternatives have been offered (see Wallis (1972), Vinod (1973) and Webb (1973)). Each of these authors considers p=4 and suggests simply generalising an existing AR(1) test by summing residuals four periods apart, rather than one period apart. Using this approach it is very straightforward to construct an exact test for serial independence against simple AR(p) errors for any chosen value of p. Some practitioners use a series of such tests in order to check for general serial independence of regression

[7] A simple AR(p) process has $u_t = \theta u_{t-p} + \varepsilon_t$ for some parameter θ.

errors.

These exact tests are, in general, only optimal when the alternative is correctly formulated. New research reported in chapter 7 below suggests that tests for simple processes have less power against the parameter which forms the alternative hypothesis when other parameters are also present. This suggests that greater effort should be directed towards generalising Kadiyala's work to an arbitrary number of parameters.

In the absence of an exact test for general p^{th} order autocorrelation, many applied workers use either a series of LM tests or a "portmanteau" test such as those suggested by Box and Pierce (1970) or Ljung and Box (1978). The preference of the LM test for this purpose, rather than the equivalent LR or Wald tests, is due to the simplicity of estimation of the null model relative to estimation of the alternative model, which requires the solution of a set of non-linear first-order conditions. The LM test is also useful in that only the maximum order of the alternative model is required; it is not necessary to specify whether the process is AR, MA or ARMA. To see this consider model (1) with $u_t = \varepsilon_t + \alpha(L)\varepsilon_t$, $\varepsilon_t \sim NID(0,\sigma^2)$, where L is the usual lag operator, $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \ldots + \alpha_g L^g$, and $\alpha = (\alpha_1 \ \alpha_2 \ldots \alpha_g)'$. This is an MA(g) model for u and, combining both equations, we can write the whole model as:

$$u_t = \left[1 + \alpha(L)\right]^{-1}(y_t - x_t'\beta).$$

Now notice that, for the purposes of constructing an LM test of $H_0:\alpha=0$, the choice of the alternative hypothesis affects the test statistic through the elements of the vector $\partial\varepsilon/\partial\alpha$ evaluated under the null hypothesis. The i^{th} element of this vector is given by

$$\partial\varepsilon_t/\partial\alpha_i = -\left[1 + \alpha(L)\right]^{-2}L^i(y_t - x_t'\beta)$$
$$= -L^i(y_t - x_t'\beta) \quad \text{under the null hypothesis.}$$

If we can find an alternative representation of ε, denoted ε* which has the property:

$$\left.\frac{\partial\varepsilon_t}{\partial\alpha_i}\right|_{\alpha=0} = \left.\frac{\partial\varepsilon_t^*}{\partial\alpha_i}\right|_{\alpha=0}$$

then it is clear that this alternative ε^* is a locally equivalent alternative to ε. The obvious example is the AR(g) model:

$$\varepsilon_t^* = \left[1 - \alpha(L)\right](y_t - x_t'\beta) \ .$$

This means that the LM test for serial independence is identical for AR(g) and MA(g) alternatives. It can similarly be shown that an ARMA(p,q) alternative is locally equivalent to an AR(j), j=max(p,q) alternative. Hence a series[8] of LM tests can potentially reveal the order, but not the form, of any autocorrelation process.

This is also the case for the "portmanteau" tests, which

[8] See Judge *et al.* (1980) for a suggested sequence of nested tests.

were originally proposed as measures of the goodness of fit in a regression model. Box and Pierce (1970) considered the properties of the sample autocorrelations of the OLS residuals

$$\hat{r}_k = \sum_{t=k+1}^{n} \hat{u}_t \hat{u}_{t-k} \Big/ \sum_{t=1}^{n} \hat{u}_t^2 .$$

They used results from Anderson (1942) to conclude that if the true disturbances were observable and had autocorrelations denoted r_k then the statistic

$$Q(r) = n(n+2) \sum_{k=1}^{p} (n-k)^{-1} r_k^2$$

would be asymptotically distributed as χ_p^2 under the null hypothesis that all autocorrelations up to lag p are jointly zero. Box and Pierce approximated $\mathrm{var}(r_k) = (n-k)/(n^2 + 2n)$ by $1/n$ and showed that

$$Q(\hat{r}) = n \sum_{k=1}^{p} \hat{r}_k^2$$

is asymptotically distributed as χ_p^2 under the null. The power of this test was questioned by several authors including Ljung and Box (1978) who recommended modifying the statistic to allow a closer approximation to the variances of the sample autocorrelations. They suggested that use of the following statistic:

$$Q^m(\hat{r}) = n(n+2) \sum_{k=1}^{p} (n-k)^{-1} \hat{r}_k^2$$

would provide a superior test although with the same asymptotic distribution. The research reported in chapter 8 below considers the properties of $Q(\hat{r})$, $Q^m(\hat{r})$ and the LM test in the context of a misspecified regression model.

This section has attempted to survey the major contributions to the literature on testing for serially independent regression disturbances. We have concentrated on correctly specified linear models with non-stochastic regressors and tried to emphasise tests which are in common use by applied researchers. This material forms the background for the new research reported later in this thesis, where we relax the assumption of correct model specification in various ways.

3.2 Heteroscedasticity Tests

In this section we discuss the literature concerned with the detection of heteroscedasticity in the disturbances of the linear regression model.

When heteroscedasticity is present in the errors of a regression model the OLS estimators of the parameter vector are inefficient, the corresponding variance estimator is biased and hence the assumed sampling distribution of the t and F statistics is incorrect. The efficient estimator, provided the other classical assumptions hold, is the GLS estimator:

$$\tilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \quad \text{where } \Omega = E(uu') = \text{diag}(\omega_{ii}).$$

There have been many tests designed to detect the presence of heteroscedasticity and in this section we attempt to survey the important offerings and highlight relevant technical details.

The coverage is by no means exhaustive but aims to include the most popular tests and those with superior properties. The discussion will be split into two subsections, dealing with the traditional formulations of the problem first, followed by a survey of the existing solutions to the problem of detecting conditional heteroscedasticity.

3.2.1 Unconditional Heteroscedasticity

If the practitioner is willing to specify the alternative hypothesis precisely then, particularly in large samples, an optimal test can be constructed using maximum likelihood based approaches. Judge *et al.* (1980) provide a thorough discussion of such procedures.

The more likely scenario is that the alternative to homoscedasticity is kept relatively broad, and in this case many tests are available. To cover the major offerings it is convenient to group similar tests together.

A popular procedure involves grouping sub-sets of observations, the best known member of this class being the Goldfeld-Quandt (1965) test. This assumes that the observations can be ordered such that $\sigma_t^2 \geq \sigma_{t-1}^2$ for all t and involves omitting c central ovservations and running separate regressions on the first and last subsamples. Under the (homoscedastic) null the ratio of the sum of squared residuals from the last (s_2) and first (s_1) subsamples is

$$r = \frac{S_2}{S_1} \sim F_{(T-c-2k)/2\ ,\ (T-c-2k)/2}\ .$$

Several variations of this procedure have been proposed. Harrison and McCabe (1979) generalised the selection of observations used in the subsamples while Theil (1971) and Harvey and Phillips (1974) suggested using the Goldfeld-Quandt procedure with BLUS and recursive residuals respectively. These tests were shown by Szröeter (1978) to be members of a wider class which includes a bounds test based on the Durbin-Watson (1951) test tables.

Notice that these tests do not specifically require any assumptions on the form of the variance function. This distinguishes them from the tests discussed below and is both a strength, due to reduced opportunity for misspecifying the alternative, and a weakness (because of the lack of information provided by rejection of the null).

Glejser (1969) considered the following model:

$$|\hat{u}_t| = z_t'\alpha + w_t$$

where $z_t' = (z_{t1},\ldots,z_{tm})$ are known constants which are possibly functions of the regressors, $\alpha' = (\alpha_1,\ldots,\alpha_m)$ is an unknown parameter vector, and w_t is a random disturbance term. He suggested using an F test of $H_0{:}\alpha=(0,..,0)'$ to test for heteroscedasticity. An alternative formulation, which uses $\hat{u}_t^2$ as the dependent variable, was proposed by Goldfeld and Quandt (1972).

If one prefers the following formulation of the variance function:

$$\sigma_t = \sigma\left\{1 + \theta g(z_t'\alpha) + o(\theta)\right\}, \qquad t=1,2,\ldots,n$$

then the Bickel (1978) test may be appropriate. Here σ and θ are unknown scalars, the z_t are as defined above, the m parameters in the vector α are assumed known, as is the function g, and $o(\theta)/\theta$ tends to zero as $\theta \to 0$. For testing the homoscedastic null, $\theta=0$, against a one-sided alternative, Bickel derived the LMP test for the case when the true disturbances are observed and then showed that this property is preserved when using consistent estimates of u in large samples. The test is very demanding informationally, however, requiring the specification of f, the density function of u_t/σ, as well as g, z_t and α.

Relaxing the assumption of a known α, Breusch and Pagan (1979) derived the LM test of H_0: $\alpha_2 = \ldots = \alpha_m = 0$ under normality. They showed that

$$LM = \frac{1}{2}\left[\sum_{t=1}^{n} z_t f_t\right]'\left[\sum_{t=1}^{n} z_t' z_t\right]^{-1}\left[\sum_{t=1}^{n} z_t f_t\right],$$

is asymptotically distributed as a χ^2 variate with m-1 degrees of freedom. Here $f_t = (\hat{u}_t^2/\hat{\sigma}^2)-1 = e_t-1$, and $\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\hat{u}_t^2$ so that the LM statistic can be alternatively obtained as one half of the explained sum of squares in a regression of e_t on z_t.

A further parametric test, due to White (1980), provides one solution to the problem of selecting the variables to use in z_t. White proposed running the following regression:

$$\hat{u}_t^2 = \alpha_0 + \sum_{i=1}^{k} \sum_{j=1}^{k} \alpha_s x_{ti} x_{tj} + \omega_t ,$$

and testing H_0: $\alpha_1 = \ldots = \alpha_{k(k+1)/2} = 0$ by using nR^2 from this regression. This statistic is asymptotically χ^2 with $k(k+1)/2$ degrees of freedom under the null hypothesis. In common with several other authors in this literature, White emphasises the importance of the maintained assumptions which underlie his test. The rejection of the null in White's test could indicate that heteroscedasticity is present but this need not be the case. Regressors which are correlated with u could lead to rejection, as could an incorrect functional form.

A class of non-parametric tests was introduced by Ali and Giaccotto (1984). Interpreting heteroscedasticity as either a location or a scale shift in the distribution of the u_t's, Ali and Giaccotto adapted results from Hájek and Šidák (1967) to establish a LMP rank test for each of these alternatives.

To conclude this section we mention the point-optimal tests suggested by Dubbelman (1978) and Evans and King (1985b,1988). These tests parameterise the variance function in a manner which allows the selection of "mid-range" parameter values. The inverse of the covariance matrix corresponding to the chosen value is then used in a Kadiyala (1970) statistic to provide a LBI test.

3.2.2 Conditional Heteroscedasticity

To distinguish conditional heteroscedasticity from its unconditional counterpart it is useful to consider a time series of observations on a suitable economic variable, such as the daily closing quotes on a particular stock or an index of many stocks. Such time series typically exhibit a temporal clustering of highly volatile prices interspersed with relatively smooth periods. The variance of the prices may well be constant over the medium term (say several months) but changing by the week. This means that our ability to forecast the variance of such speculative prices, and hence the risk inherent in holding the assets, changes over time.

Although careful researchers from Mandelbrot (1963) onwards were aware of this difficulty it was only with the publication of Engle's (1982) paper that practitioners had a formal model which captured this phenomenon and allowed improved variance forecasting. Engle proposed an Auto-Regressive model for Conditional Heteroscedasticity, known as ARCH and written as (1) with the following specification for u:

$$u_t | \psi_{t-1} \sim N(0, h_t)$$

where $$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_q u_{t-q}^2 ,$$

with $\alpha_0 > 0$ and $\alpha_i \geq 0$, $(i=1,\ldots,q)$ ensuring positive variance.

In the first ARCH application, Engle (1982) found that a large value of q was required and to reduce the computational burden he imposed a linearly declining lag structure on h_t which required the estimation of only two free parameters. Bollerslev (1986) avoided this rather restrictive formulation by generalising h_t to the following form:

$$h_t = \alpha_0 + \ldots + \alpha_q u_{t-q}^2 + \beta_1 h_{t-1} + \ldots + \beta_p h_{t-p}.$$

For positive variances in this Generalised ARCH, or GARCH(p,q) model it is not necessary that all parameters be non-negative (see Nelson and Cao (1992)), the less restrictive condition that the parameters in the infinite ARCH representation be positive is sufficient. Because the ARCH model is nested within the class of GARCH models, we shall refer to the class of conditional variance models which includes both ARCH and GARCH (but not the more general specifications of Tsay (1987)) as GARCH models.

Despite their relatively short existence GARCH models have been used in literally hundreds of applications, primarily with financial asset data but also in models of macroeconomic phenomena such as inflation. Rather less attention has been paid to the unique estimation and testing environment generated by such models. A comprehensive survey of the theoretical and applied literature was compiled by Bollerslev, Chou and Kroner (1992) while a very accessible account of recent theoretical developments is provided by Bera and Higgins (1993).

Although OLS estimation of an ARCH model still provides the best linear unbiased estimate, the non-linear Maximum Likelihood (ML) estimator is more efficient. Testing for the presence of ARCH errors concentrates on LM based procedures which use (OLS) estimates from the restricted model.

Engle (1982) derived a test for ARCH(q) errors which rejects $H_0:\alpha_1=\alpha_2=\ldots=\alpha_q=0$ for large values of

$$LM = \delta' W(W'W)^{-1}W'\delta/2 \qquad \text{where}$$

$$W' = (w_{q+1}\cdot\ldots\cdot w_n) \ ,$$

$$w_t' = (1,\hat{u}_{t-1}^2,\ldots,\hat{u}_{t-q}^2) \ ,$$

$$\delta' = \left[\frac{\hat{u}_{q+1}^2}{\hat{\sigma}^2}-1, \ \ldots \ , \ \frac{\hat{u}_n^2}{\hat{\sigma}^2}-1\right] ,$$

and $\hat{\sigma}^2$ is the ML estimator of σ^2 under H_0. An asymptotically equivalent test statistic can also be calculated as (n-q) times the R^2 from a regression of $\hat{u}_t^2$ on an intercept and q lags of itself. Both statistics are asymptotically $\chi^2_{(q)}$ under the null hypothesis.

This LM test for white noise errors against an ARCH(q) alternative is identical to the LM test against GARCH(p,q) errors (Lee (1991)). The GARCH(p,q) model can be written as an ARMA(p,q) model in the squared residuals so (using the discussion in section 3.1 above) the ARCH(q) and GARCH(p,q) models are locally equivalent alternatives to a GARCH(0,0) specification. Although the ARCH parameters are not identified under the null and the relevant block of the information matrix is singular, Lee showed that the use of any generalised inverse of I is valid so that the test exists.

In an effort to improve on the power of the LM test, Engle Hendry and Trumble (1985) suggest a one sided version which rejects for large values of the asymptotically standard normal statistic $z(LM) = \text{sign}(\hat{\alpha}_1)(nR^2)^{1/2}$. Exploitation of the one sided nature of the problem also led Lee and King (1991) to develop a locally best score test which generalises the z(LM) test to higher order processes and can be applied against either ARCH or GARCH alternatives. A third one sided test, due to Demos and Sentana (1991) uses the sum of squared t-ratios in the auxillary regression of $\hat{u}_t^2$ on an intercept and q lags of itself. This statistic is shown to be distributed as a 50:50 mixture of χ_0^2 and χ_1^2 variates and significance points of this distribution are provided by Demos and Sentana.

For the ARCH(q) model the non-negativity restrictions on the α_i parameters means that the true parameter values often lie at the boundary of the parameter space under the null hypothesis. This violates one of the regularity conditions needed for an asymptotic χ^2 distribution of the LR and Wald test statistics, and is the reason for the concentration of interest on the LM test for the detection of ARCH processes. The Nelson and Cao (1992) results suggest that this boundary problem may not apply in higher order GARCH models which leads Demos and Sentana (1991) to conjecture that the LR and Wald tests may be suitable for testing the null of homoscedasticity against GARCH(p,q) errors.

3.3 Conclusion

The aim of this chapter was to describe two major areas of econometric hypothesis testing. We have discussed the reasons for wanting to detect the presence of autocorrelation and heteroscedasticity in the errors of a regression model, and the most important techniques which have been devised for doing so. In the case of heteroscedasticity, the discussion falls naturally into two sections, concerned with conditional and unconditional heteroscedasticity. Testing for the presence of GARCH processes is currently a very active line of research which can be expected to develop rapidly over the next few years. Current work on testing against autocorrelation alternatives is concentrated either on generalising the conditions for tests[9] or on very specialised conditions[10]. A potentially fruitful avenue of future research is the attempted solution to the problem of constructing a powerful procedure for detecting general AR(p) processes.

Having surveyed the literature on testing for a scalar covariance matrix against autocorrelation and heteroscedasticity alternatives, the next chapter will address the central theme of this thesis by concentrating on tests for serial independence and relaxing the assumption of a correctly specified model.

---

[9] See Cumby and Huizinga (1992) on testing in models estimated by the Generalised Method of Moments for example.

[10] The random coefficient specification considered by Brooks (1992) is an example.

CHAPTER 4

TESTING FOR SERIAL INDEPENDENCE IN MIS-SPECIFIED LINEAR REGRESSION MODELS

In this chapter we relax the assumption of a correctly specified regression model. The existing literature which considers the properties of tests against serial correlation alternatives in mis-specified models will be surveyed and the topics which require further work will be noted. Following a brief introduction, five sub-sections will consider mis-specification due to non-normality of the errors, omitted variables, stochastic regressors, incorrect alternatives and heteroscedastic disturbances respectively. An important area of research which is not covered here concerns the pre-test issues involved in any sequential testing strategy. For analysis of the cases in which tests against autocorrelation alternatives are included in the testing strategy see King and Giles (1984), Giles and Beattie (1987) and Giles and Lieberman (1992).

We begin by defining exactly what is meant by the mis-specification of a regression model. A model is mis-specified unless **all** of the classical assumptions are valid. This very demanding criterion means that, to avoid mis-specification, the regression model must have the following properties:

Normally distributed errors

No omitted variables

Correct functional form

Zero correlation between regressors and errors

Serial independence of errors

Homoscedastic errors .

Practising econometricians invariably work, at least in the initial stages of a modeling exercise, with models which are mis-specified. The outcome of tests of specific hypotheses can lead to the estimation of a revised version of the model or the confirmation of the existing version. In either case the application of an hypothesis test is intended to provide a basis for more precise parameter estimates and more reliable inference. The question of whether these qualitative improvements actually occur, in commonly arising practical applications, is addressed in the literature on testing in mis-specified models. We now survey the existing literature on the properties of autocorrelation tests under mis-specification.

4.1 Non-normal Errors

It has been accepted for some time that many econometric models have disturbances which are not normally distributed. The most obvious example is in models using financial data, which are known to be more leptokurtic than the normal distribution, due to a relatively larger number of outliers (see Mandelbrot (1963) for example). It is therefore of practical importance to discover the consequences of non-normality for the standard econometric techniques.

Several authors have studied the properties of tests against serial correlation alternatives in models with non-normal errors. Unfortunately, most studies are restricted to the models in which the null hypothesis is true, thus avoiding potentially interesting interactions between the alternative model and non-normal errors.

One exception is a very general result due to King (1979) and applies to any statistic which is invariant to the scale of the true disturbances (the u vector of (3.1)). King showed that any such statistic has the same distribution when $u \sim N(0,\sigma^2\Omega)$ as it does when u follows any elliptically symmetric distribution with characteristic matrix[1] Ω. This result applies to any statistic which can be written as a ratio of quadratic forms in a normal random variable, a class which includes all of the popular exact tests for AR(1) errors. Because it concerns the entire distribution of any relevant statistic, the result implies that these tests have the same size, power and optimality properties when the distribution of the ε_t of (3.2) is widened to the

[1] A vector, z, follows an elliptically symmetric distribution with characteristic matrix Σ if the distribution of $\Sigma^{-1/2}z$ is invariant to orthogonal transformations (*i.e.*, if $\Sigma^{-1/2}z$ is spherically symmetric).

spherically symmetric class.

Also considering both the null and alternative models, Gastwirth and Selwyn (1980) used a Monte Carlo experiment to compare the DW test and the sign change (SC) test. They used symmetric Pareto and double exponential distributions and found the DW test to be more robust and more powerful than the SC test. Unfortunately, the model employed by Gastwirth and Selwyn was very restrictive, comprising a constant as the only regressor and this factor limits the general applicability of their results.

Several authors have considered the size of autocorrelation tests when non-normal errors are a feature of the regression model. A thorough study by Bartels and Goodhew (1981) used three design matrices and three significance levels. They found the size of the DW test against $H_1:\rho>0$ to be reasonably robust to a variety of non-normal distributions at the five per cent level but with greater than nominal size at lower levels. The direction of size distortion found by Bartels and Goodhew conflicts with a result from Smith (1976) who found that rejection frequencies were below nominal levels for those design matrices for which the DW test was not robust. Knight (1983) found the DW test to be robust to several non-normal distributions across a range of data sets. The data was, however, an important determinant of robustness when the disturbances were drawn from mixtures of normal distributions. The DW and LM tests for AR(1) errors were found by Bera and Jarque (1981) to be robust to non-normal disturbances, while Evans (1992) found the sizes of a variety of

AR(1) tests to be robust to moderately non-normal disturbances. Recently, Ali and Sharma (1993) studied the first four central moments of the null distribution of the DW test under non-normality. They found that the variance is larger (smaller) if the error distribution is long (short) tailed, the third moment is reduced by skewed distributions but the first and fourth moments are largely unaffected by non-normality.

There are several questions which remain unanswered in the literature on testing for autocorrelation in the presence of non-normal errors. It would, for example, be useful to study the effect of non-normality on the sample autocorrelation and partial autocorrelation functions. This would give an insight into the properties of the Box-Pierce and Ljung-Box tests under such conditions. A further practical benefit of this research would be in revealing the effect of non-normality on the commonly used procedure for establishing the degree of augmentation in an augmented Dickey-Fuller test for the presence of a unit root (see Dickey and Fuller (1981) for example). A further line of valuable research in this area could complete the work which has already been undertaken by restricting attention to exact AR(1) tests but considering the whole power function, rather than just the single point at which the null is true.

4.2 Omitted Variables

The omission of relevant variables is arguably the most common form of regression model mis-specification and arises as a consequence either of ignorance about the true relationship or of a lack of data for the appropriate variable. When the model is "underfitted", meaning that relevant variables are omitted, the OLS estimator of the coefficient vector is biased and inconsistent[2]. Underfitting is a more serious problem than the converse mis-specification where extraneous regressors are included and the model is said to be "overfitted". In this case the OLS estimator of the coefficient vector is inefficient but remains unbiased.

A further reason to concern ourselves with the problem of mis-specification due to omitted variables lies in the direct relationship between this problem and that of incorrect functional form. If we think of a linear regression model as being a first order truncation of a Taylor series representation of some non-linear model then it is clear that the linear model has omitted relevant variables, corresponding to the higher order terms.

Despite the likelihood, and the serious consequences, of this form of mis-specification, very little attention has been paid to the problem of autocorrelation testing in such models. The only known work is by Small, Giles and White (1993) who consider the power functions of the major exact tests for AR(1) errors in a variety of different models. They show that the rejection probability in this case is the probability that a sum

[2] This assumes that there is a non-zero correlation between the included and omitted variables.

of independent but non-central chi-square variates with one degree of freedom is negative. This contrasts with the standard case which involves central chi-squares. The noncentrality parameters are all zero, however, in the special case of linear dependence between the included and omitted variables. A further interesting theoretical finding is that the distribution of the test statistic is not independent of the scale of the error variance when relevant regressors are omitted. In their numerical evaluations, Small *et.al.* find that the true size of tests against positive AR(1) alternatives is markedly lower when seasonal intercept shift dummies, or orthogonal regressors are omitted. This result is, however, reversed when the omitted variable is a linear trend.

Further work is clearly needed on this topic. It seems to be impossible to place bounds on the power distortion of these tests caused by the omission of variables. For the reasons mentioned in the previous section, however, the effect on the sample autocorrelation and partial autocorrelation functions should be investigated.

4.3 Stochastic Regressors

When a regression model includes random variables as regressors, and these variables are also correlated with the disturbances, it is well known that OLS is an inconsistent (and therefore grossly inappropriate) estimator. Such a situation arises in a model with autocorrelated errors whenever a lagged dependent variable is also present and in this section we follow the literature in concentrating on this model. In introducing their DW test, Durbin and Watson (1950) explicitly warned against its application in models containing a lagged dependent variable, a caution which also applies to the other exact AR(1) tests discussed in the previous chapter.

Several procedures have been suggested for testing serial independence against AR(1) alternatives in models containing lagged dependent variables. The standard LM test is valid in such models as are two procedures due to Durbin (1970). The major problem with implementing the DW test is that the lagged dependent variable coefficient affects the null distribution of the test statistic, so that exact critical values cannot be obtained. Durbin's tests were aimed at adjusting the statistic to yield an asymptotic standard normal variate (known as the h statistic) or, equivalently, an asymptotic t test, under the null hypothesis. Durbin, recognising that in some models the h test cannot be applied, recommended that the t test should be used as an alternative in these cases. It is also interesting to note that Breusch (1978) found that the h test and a particular version of the t test are asymptotically equivalent to the LM test.

While Durbin's (1970) tests are not of direct relevance to the topic under study here, they have provided the motivation for several papers which study the properties of the DW test in models containing lagged dependent variables. Initial work by Nerlove and Wallis (1966) was prompted by the inappropriate use of the DW test in lagged dependent variable models. They showed that, irrespective of the validity of $H_0:\rho=0$ the DW statistic is asymptotically closer to two when the estimated model is

$$y_t = \alpha y_{t-1} + u_t \qquad t=1,\ldots,T$$

compared with the situation when the "true" residuals are used (in which case $\text{plim}(d) = 2(1-\rho)$). Despite this asymptotic bias in favour of the null, subsequent studies by Kenkel (1974, 1975) showed that in some models the DW test[3] has more reliable size and higher power than Durbin's *h* test. The implication from Kenkel's work is that the DW test is robust to the presence of a lagged dependent variable. This conclusion was disputed by Park (1975) who used the tabulated lower bound as a critical value. Both authors, however, agree that the corrections inherent in Durbin's *h* test seriously weaken its power[4].

In a more recent consideration of this issue Inder (1986) noted that power comparisons across tests are only valid when the probabilities of a type I error are standardised. This can only occur with reasonably accurate critical values. Inder considered the following model:

---

[3] Kenkel used the tabulated upper bound as a critical value

[4] Further evidence from Spencer (1975) shows that the finite sample size of the *h* test can differ substantially from that based on an asymptotic standard normal distribution.

$$y_t = \alpha y_{t-1} + \sum_{j=1}^{k} \beta_j x_{jt} + u_t \qquad t=1,\ldots,T$$

where $u_t = \rho u_{t-1} + \varepsilon_t$; $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$. Using small disturbance asymptotics he showed that the appropriate DW critical value is that from the above model with y_{t-1} omitted. A Monte Carlo experiment confirmed that, with critical values calculated in this manner, the DW test has reliable size and good power. Inder's technique can also be applied to the other exact AR(1) tests decribed above but is limited to models in which only one lagged value of y_t is present.

To summarise, it is clear that, although the standard exact tests against AR(1) errors are strictly invalid in the presence of stochastic regressors (of the lagged dependent variable type), they can have good power in such models provided that an accurate critical value is available. For models with only one such regressor the Inder (1986) method can give good results. In any case the LM test is valid but with unknown finite sample properties[5].

4.4 Incorrect Alternatives

In this section we consider the consequences of incorrectly

[5]Note that the *asymptotic* equivalence between the LM test and the *h* and *t* tests does not necessarily imply a finite sample equivalence.

formulating the alternative hypothesis in a test for autocorrelation. Suppose, for example that the true error process is a moving average of some order, but that the researcher conducts a test for autoregressive errors. In this case we should recall the discussion of locally equivalent alternatives in section 3.1 which showed that the LM test against AR(g) errors is the same as the LM test against MA(g) errors. While emphasising the need for cautious interpretation of the outcome of an LM test this result should also alert us to similar possibilities with other tests.

Given the almost routine application of exact AR(1) tests it natural that these tests have been the focus of attempts to evaluate the consequences of using an incorrect alternative hypothesis in testing for autocorrelation. The earliest work on this topic appears to be that of Blattberg (1973) who considered the power of the DW test when the true disturbances were alternatively MA(1) and AR(2). Blattberg used exact techniques and concluded that the DW test has good power against either of these processes. Furthermore, these alternatives can produce higher DW power than the AR(1) process. Blattberg also found the power to be positively related to the size of the first-order correlation coefficient. Blattberg's conclusions were supported by two later papers which used the Monte Carlo method in relatively undiscriminating[6] models (see Smith (1976) and Weber and Monarchi (1982)).

When the dependent variable in a dynamic regression is measured with error the disturbances can comprise the sum of two independent components. The same outcome occurs in a random intercept model and in models with particular aggregation problems. Using these scenarios as motivation, Revankar (1980) analysed the properties of the DW test in a model with the following disturbances:

$$\omega_t = u_t + v_t$$

where $u_t=\rho u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$, $v_t \sim N(0,\sigma_v^2)$ and $E(u_t v_t)=0$ $\forall t$. Deriving the covariance matrix for this model, Revankar concluded that $\text{plim}(d) = 2(1-\rho\lambda)$ where $\lambda = \dfrac{\sigma_u^2}{\sigma_\omega^2}$ satisfies $0\le\lambda\le1$. The DW test is therefore biased towards the null in such a model, in a manner reminiscent of the Nerlove and Wallis (1966) result. King (1982) took exception to this statement, claiming that there is no bias in the test under the null hypothesis and that the DW test is approximately LBI in this model. King went on to point out that detecting AR(1) errors is made more difficult by the presence of additional white noise, but that this does not make the DW test

[6] The design matrix used by Smith (1976) contained only a constant and a linear trend, a model for which the DW test is approximately UMPI, and therefore likely to perform extremely well.

inappropriate for such a model unless a superior test can be identified[7].

Clearly, under the null, ρ=0, so that plim(d)=2 irrespective of the value of λ. This does not, however, weaken Revankar's result that the power of the DW test is reduced by the presence of the white noise component. On the contrary, the lack of size distortion allows direct comparison of the properties of the DW test in models with and without v_t, strengthening Revankar's claim. The fact that the DW test is nevertheless approximately LBI in this model is easily explained by recalling that this criterion restricts attention to the power of tests within a small neighbourhood about the null. The model discussed by Revankar and King is similar to one employed in chapter 6, where further comment will be made.

When the true error process is MA(1) the DW test performs very well and is approximately LBI, while the Alternative DW test (ADW) is truely LBI (King (1983)). Furthermore, King and Evans (1986) showed that the DW test is also approximately LBI against a range of ARMA processes including ARMA(1,1), sums of independent ARMA(1,1) terms and a particular class of ARMA(2,1) processes.

It is clear from the foregoing discussion that the rejection of the null hypothesis in a test against AR(1) errors is (although necessary) insufficient grounds for adopting a GLS

---

[7] King reported small sample power values (for the DW test only) to support this.

estimator which assumes that AR(1) errors are present. The DW test, in particular, has been shown to have significant power against a variey of other processes. Similarly, the LM test against AR(g) errors is identical to the LM test against MA(g) errors. There is, of course, another risk which arises in mis-specified models: the AR(1) test may incorrectly fail to reject the null hypothesis when other parameters also enter the process. In the original work reported in chapter 7 below, we analyse models in which the popular exact AR(1) tests are indeed shown to have low power (relative to that attained in a correctly specified model) to detect AR(1) components of more complex processes.

4.5 Heteroscedasticity

The aim of this section is to survey the existing literature on the properties of autocorrelation tests when the regression disturbances are heteroscedastic. We consider both unconditional and conditional heteroscedasticity and discuss the literature in some detail because of the small number of papers and their relevance to the original work reported in chapters 5, 6 and 8.

There are three existing papers which examine the effect that unconditional heteroscedasticity has on the power of autocorrelation tests. In the light of the discussion above it should be no surprise that the major emphasis is again on tests against AR(1) errors.

The first work was by Harrison and M^cCabe (1975) who used a Monte Carlo experiment to study the properties of the DW bounds test, Geary's (1970) sign change test and the beta approximation test of Henshaw (1966). The approximation to the DW critical value recommended by Durbin and Watson (1971) was used to resolve the inconclusive region problem whenever this arose. Using two design matrices, three sample sizes, six values of the autocorrelation parameter and four degrees of heteroscedasticity, Harrison and M^cCabe replicated their model errors 100 times and found that "all three tests appear to be robust in the presence of heteroscedasticity".

There are several weaknesses in this study, however, which might be expected to affect the conclusions reached. First, although there is some variety in the models used the form of heteroscedasticity is always equal to X_t^{γ}, where γ varies between zero to two. Second, the number of trials is rather lower than is employed in similar experiments elsewhere in the literature. The third, and most important, difficulty concerns the method of introducing heteroscedasticity into the model. Harrison and M^cCabe allowed the white noise component, ε_t of the AR(1) process to be heteroscedastic. It will be shown in the next chapter that it is impossible to simultaneously control the degrees of autocorrelation and heteroscedasticity in such a model. More importantly the degree of heteroscedasticity decreases as the AR parameter increases so that the experiment is more likely to indicate robustness in the most important region of the parameter space.

The major weakness of Harrison and M^cCabe's work was avoided by Epps and Epps (1977) who calculated exact[8] powers for the DW test and compared them with powers of the SC test evaluated by Monte Carlo simulation. Epps and Epps criticise Harrison and M^cCabe's choice of regressors and suggest that the DW test should be subjected to "the conditions under which it is most likely to break down". They go on to claim that these conditions are provided by using a design matrix comprising a constant and the eigenvector of the A matrix corresponding to the smallest non-zero eigenvalue. The first of these statements is not sensible and the second is not correct. A study of this type should aim not to destroy the power of a test but rather to reveal its performance under a range of plausible conditions. This is especially important when test power depends upon data conditions, as it does for the DW test, and it is therefore regrettable that Epps and Epps used only one design matrix. Furthermore, if one was to seek "the conditions under which it (the DW test) is most likely to break down" then one should certainly avoid models in which the X matrix is spanned by the eigenvectors of A[9], as the DW test is approximately UMP similar for these matrices (see section 3.1).

Epps and Epps assumed that the variance of u_t was given by

---

[8] Curiously, Epps and Epps consistently refer to the bounds test, despite their use of the exact version of the DW test.

[9] Methods for determining the extent to which the eigenvectors of A span X are discussed by Cassing and White (1983)

$\sigma_t^2=\delta+\gamma X_t^2$ and measured the degree of heteroscedasticity through the parameter λ defined as the ratio of the maximum to the minimum of σ_t^2. They employed two sample sizes and seven values of the AR parameter ranging from -0.9 to 0.9. Epps and Epps found the DW test to be quite robust to heteroscedasticity with only minor size distortion (upwards) and slight falls on power. The Monte Carlo power results reported for the SC test are not, however, comparable with the DW powers. This is because the sizes are markedly different, possibly due to insufficient replications[10]. Ignoring this problem, Epps and Epps concluded that their findings "constitute strong evidence for the superiority of the bounds test".

The third paper on this topic, due to Giles and Small (1991) examined the DW test under a range of data conditions[11]. This study used seven regressor matrices and two sample sizes, the intention being to reveal the effect that the data has on the power of the test under this type of mis-specification. Powers were evaluated using the exact distribution so that no Monte-Carlo experiment was required. Giles and Small found that when the design matrix was comprised of eigenvectors of the A matrix the power of the test was increased by heteroscedasticity through the middle of the autoregressive parameter space, with negligible distortion as ρ approached either zero or unity. Emphasising the

[10] 200 replications were used for the sample of 30 observations, giving a reported size of 4%, while 350 replications for the 15 observation sample resulted in a 1% reported size.

[11] The analysis of chapter 6 includes an extension of this work.

importance of the data, however, they found that with all other data the power of the DW test was considerably lower when heteroscedasticity was present. In some of these latter cases the power functions ceased to be monotonic in ρ and fell alarmingly as ρ approached unity. Chapter 6 will outline the major reasons for the serious decline in power found by Giles and Small (1991) and the connections between this model and one discussed in section 4.4 in which the variance has two additive components.

Many applied papers now explicitly consider variance models which are related to the autoregressive conditional heteroscedasticity (ARCH) model introduced by Engle (1982). Empirically, ARCH models have proven very useful in modelling asset returns data in particular, and financial data generally. The class of ARCH related models will be discussed more fully in chapters 5 and 8. Our concern here is to outline the literature concerned with testing for autocorrelation when the variance is conditionally heteroscedastic. This literature includes only a very small number of papers.

In a study aimed at the estimators of and tests for ARCH processes, Engle, Hendry and Trumble (1985) observed that the standard AR(1) process

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad t=1,\ldots,T$$

has the following conditional variance characteristics:

$$E(u_t^2 | u_{t-1}) = \sigma_\varepsilon^2 + \rho^2 u_{t-1}^2 .$$

Thus the autoregressive representation of the mean induces an ARCH process in the variance. However, if the true data generating process is

$$u_t = \xi_t(\gamma + \alpha u_{t-1}^2)^{1/2}, \qquad \xi_t \sim IN(0,1)$$

then the standard exact AR(1) tests are inconsistent tests of $H_0{:}\alpha=0$. Conjecturing that the DW test may nevertheless be useful in finite samples, Engle, Hendry and Trumble fitted a response surface for the test and estimated the following power function (size corrected to a 10% level)

$$\hat{P}_{DW} = \{1 + \exp(2.18 - 0.044\alpha T/(1 + 0.0005(\alpha T)^2))\}^{-1}$$

The DW test in this model therefore has a maximum power of 23% which occurs when $\alpha T=45$.

This result shows that if $\rho=0$ but $\alpha\neq0$ then the DW test has very low power, which does not necessarily increase with the sample size. The scenario considered does not, however, apply to a model in which the errors are both autoregressive and conditionally heteroscedastic. The model is mis-specified in that the errors are generated by an ARCH process rather than an AR(1) process. It is, however, reasonable to suppose that the errors are autoregressive in both the mean and the variance. The next chapter discusses models which accomodate both of these phenomena, while chapter 8 evaluates the properties of a variety of tests for autocorrelation in the context of these models.

The sample autocorrelations of a white noise time series have a variance of approximately 1/T. When ARCH effects are present, however, it can be shown that the variance of the sample autocorrelations is given by[12]

$$S(\tau) = \frac{1}{T}\left\{1 + \frac{\gamma_{u^2}(\tau)}{\sigma^4}\right\}.$$

where $\gamma_{u^2}(\tau)$ is the autocovariance of the squared process at lag τ, and σ^4 is the squared unconditional variance of the u_t's. Furthermore, the second bracketed term above is strictly greater than unity for an ARCH process, which must therefore have a greater variance than suggested by the standard approximation. This result is due to Milhöj (1985) who also commented on the need for autocorrelation tests to allow for the presence of ARCH effects.

Following this work, Diebold (1986) considered the effect of ARCH errors on the sample autocorrelations and their associated standard errors. Noting that consistent estimators for $\gamma_{u^2}(\tau)$ and σ^4 are readily available, Diebold recommended constructing confidence intervals for the autocorrelations such that

$$\rho_u(\tau) = 0 \pm 1.96 \ (\hat{S}(\tau))^{1/2}$$

where $\hat{S}(\tau)$ is $S(\tau)$ with $\gamma_{u^2}(\tau)$ and σ^4 replaced by their consistent estimates. Observing that the recommended correction relies on the existence of the fourth moment of u, which is by no means always satisfied, Diebold nevertheless claims that even if this condition is not met the best finite sample correction to

[12] In early work on this topic, Weiss (1984) used an asymptotic approximation to the variance of the sample autocorrelations.

the standard errors is still provided by the above procedure.

As a further caveat on the Diebold correction, the analysis assumes that the series u_t is directly observed, rather than being an estimate of an unobservable disturbance series. The findings of Box and Pierce (1970) are invoked to claim validity of the method when the statistics are tested against an asymptotic χ^2 distribution.

The major application of the correction is to the portmanteau tests of Box and Pierce (1970) and Ljung and Box (1978), both of which can be readily modified to incorporate the alternative expression for the variance of the autocorrelations of u. Diebold reports the results of a Monte Carlo experiment which shows that the corrections accurately correct severe over rejection problems under the null hypothesis, with both the Box-Pierce and Ljung-Box tests. The experiment is conducted on observed series, however, and a sample size of 500 observations was used. Both of these factors could be expected to favour the suggested correction. There is also a need to evaluate the power consequences of modifying the portmanteau tests in this way. Each of these questions is investigated in the study reported in chapter 8.

Several papers develop models in which autocorrelation and ARCH processes are simultaneously present. These are not relevant to the mis-specification issues considered here but are discussed in the following chapter.

Mention should be made of the work by Wooldridge (1991) who addresses exactly the mis-specification which is considered here. Rather than explore the consequences of ARCH errors for existing autocorrelation tests, Wooldridge proposes a "robust" test which is guaranteed to have the correct size asymptotically[13]. Strangely, Wooldridge makes no empirical evaluation of the properties of this test procedure. The study reported in chapter 8 includes Wooldridges test in an effort to reveal not only the finite sample characteristics but also the power properties of the technique.

To conclude this chapter we observe that, in general, considerably more effort has been directed to the construction and application of tests for autocorrelation than to the careful evaluation of their respective strengths and weaknesses. Several notable omissions have been noted above in connection with mis-specifications due to non-normal errors, omitted variables, incorrect alternative hypotheses and heteroscedasticity. The following chapters go some distance towards correcting the relative imbalance in the literature in respect of problems raised by the presence of heteroscedasticity (chapters 5,6 and 8) and the choice of an inappropriate alternative hypothesis (chapter 7).

[13] This approach is also taken by Domowitz and Hakkio (1985) who use White's (1980) covariance matrix estimator in conjunction with the standard LM test for autocorrelation.

CHAPTER 5

THE SIMULTANEOUS MODELLING OF AUTOCORRELATION AND HETEROSCEDASTICITY

The previous chapter surveyed the existing literature on the subject of testing for autocorrelation in mis-specified regression models. In almost all of the work covered in that chapter the modelling method was unambiguous. To check the effect of non-normal errors, for example, one would proceed by deriving the exact distribution of the test statistic under a more realistic distributional assumption, such as multivariate t distributed errors. In cases where this approach proved analytically intractable the Monte Carlo method could be employed. The important point is that there is no ambiguity about how such a mis-specification would enter an autoregressive regression model. This is not the case when the regression errors are assumed to be both autocorrelated and heteroscedastic.

Despite the pedagogically convenient separation of these two effects in econometrics textbooks, applied workers frequently test the same regression for both problems. This practice raises several important theoretical questions, the answers to which have implications for applied workers. There is, as the previous chapter has shown, very little evidence in the literature concerning the effect of heteroscedasticity on the power of autocorrelation tests. The converse of this problem in similarly under-researched. A third important line of research which has not been explored concerns the pre-test issues involved in this testing strategy and the subsequent estimation of the preferred model. This major focus of this thesis is the first of these questions: the effect that variance mis-specification has on the properties of autocorrelation tests. This chapter, however, addresses a problem which is common to all three avenues of research. Before we can look for answers to these questions we must have a clear idea of how to characterise a regression model which has errors which are autocorrelated and heteroscedastic. It will be shown below that there is more than one solution to this modelling task.

In what follows we shall find it convenient to separate the discussions of conditional heteroscedasticity from those concerning the more traditional unconditional heteroscedasticity. There are several reasons for this separation. Clearly, these two variance assumptions have very different effects. It is also well known that unconditional heteroscedasticity can be fully described by the appropriate covariance matrix (assuming normality), whereas the observations of a conditionally heteroscedastic series are related through their fourth moments and are therefore more difficult to specify analytically. For the purpose of this chapter there is a further reason for separation which is both more fundamental and more surprising. There is an existing literature which explicitly considers the joint modelling of conditional heteroscedasticity and autocorrelation, while the very few papers which consider the unconditional analogue pay minimal attention to the modelling issue.

5.1 Unconditional Heteroscedasticity

In this section it is assumed that the regression error variance is not constant but it is exogenous. In particular, we can think of the variance as some deterministic function of one or more of the regressors in the model. We use the following linear regression model with AR(1) errors:

(1) $$y = X\beta + u$$

(2) $$u_t = \rho u_{t-1} + \varepsilon_t \ , \qquad t=1,..,T$$

where y, u and ε are nx1 vectors of realisations of the dependent variable and two stochastic error terms respectively, $\varepsilon_t \sim IN(0,\sigma_\varepsilon^2)$, X is (nxk) and of full rank and β is a kx1 parameter vector.

If one was to conduct a simulation experiment to study the power of a test for AR(1) errors in this model, the normal procedure would be to generate the ε vector and use it to form u for a selected value of ρ. The logical extension of this technique to allow for heteroscedasticity would then simply involve constructing ε such that[1] $\varepsilon_t \sim N(0,\sigma^2_{\varepsilon_t})$. It may well be that the resulting model represents reality in some applications but as a vehicle for studying the power of a test for AR(1) errors it is seriously flawed as the following theorem shows.

Theorem 5.1

Define $h_u = \dfrac{\max \sigma^2_{u_t}}{\min \sigma^2_{u_t}}$ as a measure of the degree of heteroscedasticity of series u.

If $u_t = \rho u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim (0,\sigma^2_{\varepsilon_t})$, $\sigma^2_{\varepsilon_t} > \sigma^2_{\varepsilon_{t-1}}$ for all $t=1,..,T$, and the first sample observation is drawn from a stationary distribution, then

(a) h_u decreases with ρ for $\rho>0$ and increases with $|\rho|$ for $\rho<0$

(b) $\lim\limits_{\rho\to 1} h_u = 1$

Proof: see appendix 5.1

Part (a) of this theorem shows that a simulation trial which introduces heteroscedasticity to the model through the ε vector will not be capable of simultaneously controlling the degrees of

[1] It will be recalled from chapter 4 that this is exactly the method used by Harrison and McCabe (1975).

autocorrelation and heteroscedasticity. The implication of part (b) is that in such a model the disturbance variance approaches a constant as the autoregressive process approaches the non-stationarity boundary. The existing literature on the limiting power of AR(1) tests (see section 6.2 below), combined with the well known decline in the relative efficiency of OLS estimation as ρ increases, suggests that this is a particularly interesting and important region of the parameter space which deserves careful investigation.

The obvious corollary to Theorem 5.1 is that if such a model does represent reality then the effect of unconditional variance mis-specification on the probability of detecting positive values of ρ becomes insignificant as ρ approaches unity. It may well be, however, that test power effects are serious for more moderate degrees of autocorrelation, in which case we would expect to see mis-specified power curves converging to the properly specified curves as ρ approached unity.

In the light of Theorem 5.1 it was decided to introduce heteroscedasticity directly into the u vector in our own work. This decision still leaves at least two options for the precise form of the covariance matrix of u, as will be seen below.

Let V represent the standard (homoscedastic) covariance matrix of an AR(1) process, neglecting the scale parameter. One possible form for V when u is also heteroscedastic is constructed by superimposing the vector of error variances onto the leading diagonal of V. This is the form used by Giles and Small (1991) and will be denoted V*:

$$V^* = \begin{bmatrix} \sigma_1^2 & \rho & \rho^2 & . & . & . & \rho^{T-2} & \rho^{T-1} \\ \rho & \sigma_2^2 & \rho & . & . & . & \rho^{T-3} & \rho^{T-2} \\ \rho^2 & \rho & \sigma_3^2 & . & . & . & \rho^{T-4} & \rho^{T-3} \\ . & & & . & & & & . \\ . & & & & . & & & . \\ . & & & & & . & & . \\ . & & & & & & . & . \\ . & & & & & & . & \rho \\ \rho^{T-1} & & & & & & \rho & \sigma_T^2 \end{bmatrix} .$$

This covariance matrix arises naturally from considering AR(1) errors in the context of a Hildreth-Houck (1968) random coefficient model, for example. Although this model has all coefficients random, V* can also be derived from a model in which only the intercept is random and in this context some extra insight is gained. Suppose that

(3) $y_t = x_t'B + \mu_t + u_t$

(4) $u_t = \rho u_{t-1} + \varepsilon_t,$

where $\mu_t \sim (0,\sigma_\mu^2)$ is the random intercept and $\varepsilon_t \sim (0,\sigma_\varepsilon^2)$ is independent of μ_t. This is a particular form of the variance-components model which is often used in the analysis of panel data (see Hsiao(1986), for example). In this model the covariance matrix of $\nu_t = \mu_t + u_t$, neglecting the scale factor is given by V* with $\sigma_t^2 = (1+\sigma_\mu^2)$. It can now be easily seen that restricting u_t to have constant variance of $\sigma_t^2 = \lambda$ implies that V* reflects the following autoregressive process:

(5) $$\lambda u_t = \rho^s u_{t-s} + \varepsilon_t \; .$$

Here the first autocorrelation is ρ/λ while all subsequent ones are ρ. Notice that λ must exceed unity so that the first autocorrelation is weaker than all others. Because the exact AR(1) tests considered in chapter 6 are all invariant to the scale of the disturbance variance, for the purposes of analysing such tests, the matrix V* is formally equivalent to the matrix studied by Revankar (1980) and King (1982) and referred to in the previous chapter. It will be proven in section 6.1 that, when the true process is given by (5), the popular exact tests for H_O: ρ = 0 in (2) are seriously weakened. This result will give finite sample support to Revankar's claim of reduced power for the Durbin Watson test in the variance components model.

The V* matrix reveals a connection between the heteroscedasticity induced by random coefficients and the nature of an autoregressive process. The implications of this link for the properties of autocorrelation tests are covered in the next chapter, while a somewhat different random coefficient model arises in the next section of this chapter. A reasonable question at this stage, however, concerns the relevance of V* to applied econometricians.

Random coefficient models have been found useful in a wide variety of applications including agricultural production (Hoque (1991)), finance (Easton and Zmijewski (1989)) and energy economics (Bartels and Fiebig (1990)). The assumption has also been used to facilitate the pooling of equations in a panel data study by Liu and Tiao (1980). Furthermore, there are undoubtedly many applied studies in which the data would suggest a random coefficient model if this possibility was entertained by the researcher. The algebraic and numerical analysis of the properties of autocorrelation tests in a model with a V* covariance matrix applies equally well whether or not the researcher recognises that the coefficients are random.

An alternative method for modelling an autocorrelated error term is now introduced. Recall the fact that the correlation between two random variables is equal to their covariance divided by the product of their standard deviations. If u follows an AR(1) process then the correlation between u_t and u_{t-s} is ρ^s. Simply imposing the additional requirement that $E(u_t^2) = \sigma_t^2$ implies that the covariance matrix of u is given by:

$$V^{**} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 & \cdot & \cdot & \cdot & \rho^{T-1}\sigma_1\sigma_T \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & & & & \rho^{T-2}\sigma_2\sigma_T \\ \rho^2\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 & & & & \cdot \\ \cdot & & & \cdot & & & \cdot \\ \cdot & & & & \cdot & & \cdot \\ \cdot & & & & & \cdot & \cdot \\ \rho^{T-1}\sigma_1\sigma_T & \cdot & \cdot & \cdot & \cdot & \cdot & \sigma_T^2 \end{bmatrix}$$

The derivation of V**, sketched above, emphasises the simultaneous preservation of a first order autoregressive process and heteroscedasticity, and in this respect V** provides an important contrast to V*. As shown earlier, V* can be seen as representing an autoregressive process in which the first autocorrelation is weaker than all subsequent ones. Thus it mis-specifies the autoregressive process in a way which V** does not. Although one might be tempted to conjecture that V* would therefore have greater effect on the power of AR(1) tests than V**, such speculation is only partially correct, as will be shown in chapter 6.

In their earlier work on the Durbin-Watson test Epps and Epps (1977) were not specific about their assumed true covariance matrix. Attempts to replicate their results suggest that V** was the matrix used but some doubt remains over this question as it was not possible to obtain exact agreement with their published results. The discrepancies are believed to be due to the different algorithms which were used, firstly to extract the eigenvector which Epps and Epps used as the regressor in their study, and secondly in the computation of the rejection probabilities themselves.

5.2 Conditional Heteroscedasticity

The analysis and discussion up to this point has followed the traditional approach to the non-constant variance phenomenon. In this paradigm heteroscedasticity is a secondary nuisance which must be accomodated in some way in the interests of estimation efficiency. The standard techniques have involved assuming that the error variance is proportional to some (not necessarily linear) function of a subset of the explanatory variables in the regression model. Several of these formulations will be introduced in chapter 6. We now consider some of the more recent developments in the modelling of heteroscedasticity and, in particular, models which accomodate both autocorrelation and conditional heteroscedasticity.

For several decades it has been known that the risk associated with the holding of assets (particularly financial assets) varies over time. This risk is measured by the variance of the asset price in the appropriate market. In particular it was observed that the variances of a given asset were similar in magnitude for time periods that were close together, a phenomenon known as volatility clustering. A further empirical fact is that the probability distributions of such series were observed to have more kurtosis than the normal distribution due to relatively greater proportions of outlier values. These empirical realities have important implications for asset pricing models, which generally describe the value of an asset in terms of the expected return to holding it deflated by the associated capital risk. If the risk varies over time then a rational investor must forecast the risk in order to value the asset. The best forecast is one which evolves in response to new information; it is said to be conditional on the available data.

These ideas provided the motivation for the work of Engle

(1982) who introduced the following autoregressive conditional heteroscedasticity, or ARCH, model for the variance of the series, u_t $t=1,\ldots,T$:

$$(6) \qquad \sigma_t^2 = \sigma^2 + \sum_{i=1}^{q} \alpha_i \, u_{t-i}^2$$

where $\sigma^2 > 0$, $\alpha_i \geq 0$ for all i and $\sum \alpha_i < 1$. Initial applications of the ARCH model imposed a linearly declining lag structure on the coefficients in order to give more weight to the most recent information. The linear ARCH model above was rapidly generalised in several directions. Bollerslev (1986) allowed for direct feedback of lagged variance terms in his generalised ARCH or GARCH model. The GARCH(p,q) model is written as:

$$(7) \qquad \sigma_t^2 = \sigma^2 + \sum_{i=1}^{q} \alpha_i \, u_{t-i}^2 + \sum_{j=1}^{p} B_j \, \sigma_{t-j}^2 .$$

This model can be readily reinterpreted as an ARMA representation of the squared process, u_t^2, an approach which allowed Bollerslev (1988) to suggest appropriate techniques for the identification of the orders of p and q.

The analysis of this chapter, and of chapter 8, concerns only the above formulations of conditional heteroscedasticity. We therefore exclude further generalisations of the ARCH model, such as integrated and exponential GARCH models and the GARCH in mean (GARCH-M) model. For a recent survey of the literature on these, and other ARCH related processes, see Bollerslev, Chou and Kroner (1992).

The discussion to this point applies to ARCH models for a single observed series. When that series is the disturbance term in a regression model, a separate variance equation is specified in addition to the usual expression for the mean of the dependent variable. A logical extension of this approach to include possibly autocorrelated disturbances is the following model

$$(8) \qquad u_t = \sum_{j=1}^{s} \rho_j u_{t-j} + \varepsilon_t$$

$$(9) \qquad h_t = \sigma^2 + \sum_{i=1}^{q} \alpha_i \, \varepsilon_{t-i}^2$$

where h_t is the variance of ε_t conditional on the information set available at time t-1. A similar model was used by Weiss (1984) to construct univariate models for a set of US macroeconomic time series. Weiss did not use a regression model (so that u_t was the time t observation of the actual series under study) but allowed for full ARMA[2] processes in the mean of the series. He found that, with one exception, the levels of each of the 16 series could be modelled in this way; taking logarithms was less successful, however, with several series having no ARCH component in the logs.

These results suggest that many economic time series may have an ARMA-ARCH representation of the type described above. They also suggest that the inadvertent omission of such a variable from a regression model would lead to errors with these

---

[2] In fact only one of Weiss's final models actually used a full ARMA representation, the majority being either AR(1) ar MA(1).

characteristics.

The extension of Weiss's model to accomodate GARCH innovations is straightforward and requires no further discussion.

In the work reported in chapter 8 we consider the power functions of several tests for autoregressive errors when the supposedly white noise component of the AR process in fact follows a GARCH model. This work studies exact tests for both AR(1) and AR(4) errors as well as asymptotic tests such as the LM and Ljung-Box tests which are regularly used in empirical applications of GARCH models. We also evaluate the test proposed by Wooldridge (1991) and discussed in the previous chapter.

The Weiss model described above, while intuitively attractive, is not a particularly parsimonious parameterisation of autoregression in both conditional mean and conditional variance. This is because it does not parallel Engle's (1982) model in which the conditional variance is a function only of past regression errors. Weiss's h_t is additionally dependent on past values of ε_t, the innovation series.

Several authors have drawn attention to the parallels between random coefficient models and ARCH models (see Tsay (1987) and Wolff (1988) for example) and a formal link was established by Bera and Lee (1991). The White (1982) information matrix test was shown by Chesher (1984) to be equivalent to an LM test for specification error against the alternative of parameter heterogeneity. Bera and Lee (1991) show that a special case of one component of the White test (when applied to a model with AR errors) is identical to Engle's (1982) LM test for ARCH. White's test, therefore, is unable to distinguish between ARCH errors and autoregressive errors with random coefficients.

This link provides a strong justification for the class of models which parameterise ARCH models as being autoregressive with random coefficients. This class of models was proposed by Bera, Higgins and Lee (1992) (henceforth BHL) and can be written as:

$$y_t = x_t'B + u_t \tag{10}$$

$$u_t = \sum_{j=1}^{p} \phi_{jt} u_{t-j} + \varepsilon_t \tag{11}$$

$$= \sum_{j=1}^{p} (\phi_j + \eta_{jt}) u_{t-j} + \varepsilon_t \tag{12}$$

where the definition of terms in the first equation is standard. The disturbance u_t is autoregressive of order p with random parameters ϕ_{jt} which are the sum of a fixed term, ϕ_j, and a random component, η_{jt}. It is assumed that $\eta_t = (\eta_{1t},\ldots,\eta_{pt})'$ is a sequence of iid vectors such that $\eta_t \sim (0,\Sigma)$, $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_T)$ is also iid with $\varepsilon \sim (0,\sigma^2 I)$ and η_t, ε_t are mutually independent.

Defining $\phi = (\phi_1,\ldots,\phi_p)'$ and $\underline{u}_{t-1} = (u_{t-1},\ldots,u_{t-p})$ and conditioning on Ψ_{t-1}, the information set available at time t, we can write the conditional mean of u_t as

$$\mu_t = E\left[\left\{ \sum_{j=1}^{p} (\phi_j + \eta_{jt}) u_{t-j} + u_t \right\} \mid \Psi_{t-1}\right]$$

$$= \sum_{j=1}^{p} \phi_j u_{t-j} = \phi' \underline{u}_{t-1} . \tag{13}$$

Similarly, the conditional variance of u_t is given by

$$h_t = E\left[\left(\sum_{j=1}^{p} h_{jt} u_{t-j} + u_t\right)^2 \mid \Psi_{t-1}\right]$$

$$= E\left[\underline{u}'_{t-1}\eta_t\eta_t\underline{u}_{t-1} + 2\varepsilon_t\underline{u}'_{t-1}\eta_t + \varepsilon_t^2 \mid \Psi_{t-1}\right]$$

$$= \underline{u}'_{t-1}\Sigma\, \underline{u}_{t-1} + \sigma^2 . \tag{14}$$

Notice that, by imposing zero restrictions on a subset of the η_{jt}'s, this model is sufficiently flexible to allow a higher order for the AR component than the ARCH component. Thus an AR(4)-ARCH(1) model, for example, is possible. It is not, however, possible to construct the converse AR(1)-ARCH(4) model as a non-zero η_{jt} will still provide an autoregressive effect at any lag j for which $\phi_j = 0$.

An attractive feature of this model is the unrestricted nature of the conditional variance function. The linear ARCH of Engle (1982) is obtained from the BHL model when Σ is a diagonal matrix. When this special case does not apply, however, the dependence of h_t on cross products between previous errors allows for a more general specification of the conditional variance. BHL refer to this model as augmented ARCH, or AARCH. One consequence of this augmentation is that the BHL ARCH process need not be symmetric and so is capable of incorporating the leverage effects described by Nelson (1991) for example.

It is clear that the first two conditional moments of the BHL model and the standard ARCH model are identical. If the distribution of the underlying innovation series, ε_t, is assumed normal then all the moments are identical, and the two processes are therefore also identical (Bera and Higgins (1993)). The above interpretation of ARCH as a random coefficient AR model can also be extended to the GARCH model with relatively little effort. The u_t are rewritten as

$$u_t = \sum_{i=1}^{q} \phi_{it} u_{t-j} + \sum_{j=1}^{p} \delta_{jt} \sqrt{h_{t-j}}\ \varepsilon_t \tag{15}$$

where $\delta_t = (\delta_{1t}, \ldots, \delta_{pt})'$ is an iid sequence with zero mean and diagonal covariance matrix B_p and independent of ϕ_t and ε_t. The conditional mean is unchanged from the previous model and, defining $\underline{\sigma}_{t-1} = \left(\sqrt{h_{t-1}}, \ldots, \sqrt{h_{t-p}}\right)'$, the conditional variance of u_t can be written as

$$h_t = \underline{u}'_{t-1} A_q \underline{u}_{t-1} + \underline{\sigma}_{t-1} B_p \underline{\sigma}_{t-1} + \sigma^2. \tag{16}$$

When A_q is a diagonal matrix (16) is a GARCH(p,q) model but in general this is not necessary. Models in which A_q is not diagonal are referred to by BHL as generalised augmented ARCH, or GAARCH, models. In common with (14) above, restrictions on the ϕ_i's can be used to vary the orders of the GARCH and autoregressive components of the process.

5.3 Conclusion

This chapter has shown that the joint modelling of autocorrelation and unconditional heteroscedasticity can be approached in several different ways. In one of these (that used by Harrison and McCabe (1975)) the size of the first autocorrelation is a function of the degree of heteroscedasticity, so that the particular consequences of these two effects cannot be distinguished. Each of the other two models are incorporated in the analysis of the next chapter.

The combination of autocorrelation and conditional heteroscedasticity in a single model has been explicitly considered in two distinct ways by previous authors. Each of these models are used in chapter 8.

APPENDIX 5.1

Theorem 5.1

Define $h_u = \dfrac{\max \sigma^2_{u_t}}{\min \sigma^2_{u_t}}$ as a measure of the degree of heteroscedasticity of series u.

If $u_t = \rho u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim (0, \sigma^2_{\varepsilon_t})$ and $\sigma^2_{\varepsilon_t} > \sigma^2_{\varepsilon_{t-1}}$ for all $t=1,..,T$, then

(a) h_u decreases with ρ for $\rho > 0$

increases with $|\rho|$ for $\rho < 0$

(b) $\lim\limits_{\rho \to 1} h_u = 1$

Proof:

Begin by defining the following matrix

$$
T = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & . & . & . & 0 \\ -\rho & 1 & 0 & & & . \\ 0 & -\rho & 1 & & & . \\ . & & & . & & . \\ . & & & & . & 0 \\ 0 & & . & . & 0 & -\rho & 1 \end{bmatrix}
$$

which is non-singular with the property $Tu = \varepsilon$. Hence $u = T^{-1}\varepsilon$ and $E(uu') = \Sigma = T^{-1}\Omega T^{-1}$, where $\Omega = \mathrm{diag}\ (\sigma^2_{\varepsilon t})$. Expansion of the quadratic form reveals that the diagonal elements of Σ (which are the variances of u) are given by

$$
\sigma^2_u = \left[\frac{\sigma^2_{\varepsilon 1}}{1-\rho^2}\right], \left[\frac{\rho^2\sigma^2_{\varepsilon 1}}{1-\rho^2} + \sigma^2_{\varepsilon 2}\right], \left[\frac{\rho^4\sigma^2_{\varepsilon 1}}{1-\rho^2} + \rho^2\sigma^2_{\varepsilon 2} + \sigma^2_{\varepsilon 3}\right], \ldots
$$

$$
\ldots, \left[\frac{\rho^{2T-2}\sigma^2_{\varepsilon 1}}{1-\rho^2} + \rho^{2T-4}\sigma^2_{\varepsilon 2} + \ldots + \rho^2\sigma^2_{\varepsilon (T-1)} + \sigma^2_{\varepsilon T}\right].
$$

The difference between successive variances is therefore:

$$
\sigma^2_{ut} - \sigma^2_{u(t-1)} = \left[\sigma^2_{\varepsilon t} + \rho^2\sigma^2_{\varepsilon (t-1)} + \ldots + \rho^{2(t-2)}\sigma^2_{\varepsilon 2} + \frac{\rho^{2(t-1)}\sigma^2_{\varepsilon 1}}{1-\rho^2}\right]
$$

$$
- \left[\sigma^2_{\varepsilon (t-1)} + \rho^2\sigma^2_{\varepsilon (t-2)} + \ldots + \rho^{2(t-3)}\sigma^2_{\varepsilon 2} + \frac{\rho^{2(t-2)}\sigma^2_{\varepsilon 1}}{1-\rho^2}\right]
$$

$$
= \sigma^2_{\varepsilon t} + (\rho^2-1)\left[\sigma^2_{\varepsilon (t-1)} + \rho^2\sigma^2_{\varepsilon (t-2)} + \ldots + \rho^{2(t-2)}\frac{\sigma^2_{\varepsilon 1}}{1-\rho^2}\right]
$$

$$
= (\sigma^2_{\varepsilon t} - \sigma^2_{\varepsilon (t-1)}) + \rho^2(\sigma^2_{\varepsilon (t-1)} - \sigma^2_{\varepsilon (t-2)}) + \ldots
$$

$$
\ldots + \rho^{2(2t-2)}(\sigma^2_{\varepsilon 2} - \sigma^2_{\varepsilon 1})
$$

$$
> 0 \quad ; \text{ since } \sigma^2_{\varepsilon t} > \sigma^2_{\varepsilon (t-1)} \text{ for all t.}
$$

The largest σ^2_{ut} therefore occurs at t=T, the smallest at t=1, and

h_u is the ratio of these two values.

$$h_u = \frac{\sigma^2_{uT}}{\sigma^2_{u1}} = \frac{1-\rho^2}{\sigma^2_{\varepsilon 1}}\left[\frac{\rho^{2(T-1)}\sigma^2_{\varepsilon 1}}{1-\rho^2} + \rho^{2(T-2)}\sigma^2_{\varepsilon 2} + \ldots + \rho^2\sigma^2_{\varepsilon(T-1)} + \sigma^2_{\varepsilon T}\right]$$

$$= \rho^{2(T-1)} + (1-\rho^2)\left[\rho^{2(T-2)}\frac{\sigma^2_{\varepsilon 2}}{\sigma^2_{\varepsilon 1}} + \rho^{2(T-3)}\frac{\sigma^2_{\varepsilon 3}}{\sigma^2_{\varepsilon 1}} + \ldots + \frac{\sigma^2_{\varepsilon T}}{\sigma^2_{\varepsilon 1}}\right].$$

Define $R_j \equiv \dfrac{\sigma^2_{\varepsilon j}}{\sigma^2_{\varepsilon 1}}$ and note that $R_j > R_{j-1}$ for all $j = 1,\ldots,T$.

Expanding powers of ρ in the previous expression for h_u gives:

$$h_u = \rho^{2(T-1)} + R_2(\rho^{2(T-2)} - \rho^{2(T-1)}) + \ldots + R_{T-1}(\rho^2 - \rho^4) + R_T(1-\rho^2). \qquad (1)$$

We can now differentiate with respect to ρ to obtain

$$\frac{\partial h_u}{\partial \rho} = 2(T-1)\rho^{2T-3} + R_2\{2(T-2)\rho^{2T-5} - 2(T-1)\rho^{2T-3}\} + \ldots$$
$$\ldots + R_{T-1}\{2\rho - 4\rho^3\} - R_T\{2\rho\}$$
$$= 2(T-1)\rho^{2T-3}(1-R_2) + 2(T-2)\rho^{2T-5}(R_2-R_3) + \ldots$$
$$\ldots + 4\rho^3(R_{T-2} - R_{T-1}) + 2\rho(R_{T-1} - R_T). \qquad (2)$$

Observe that all the bracketted terms involving R_j's are negative in this expression so that $\dfrac{\partial h_u}{\partial \rho} < 0$ for $\rho > 0$. Furthermore only odd powers of ρ appear in (2) so that, for $\rho < 0$, each term is strictly positive and $\dfrac{\partial h_u}{\partial \rho} > 0$ in this case. This completes the proof of part (a) of the theorem.

To establish part (b), consider equation (1) above and observe that, as $\rho \to 1$, h_u collapses to a single term, $\rho^{2(T-1)}$, which approaches unity, implying homoscedasticity for the u_t's.

CHAPTER 6

TESTING AGAINST AR(1) ALTERNATIVES IN THE PRESENCE OF UNCONDITIONAL HETEROSCEDASTICITY

6.1 Introduction

In applied econometric work based on time series data, testing for serial independence of the regression disturbances is a routine and necessary practice. Such testing can reveal to the researcher signs of specification error or poor explanatory performance associated with the model, as well as information about the appropriate estimation technique to employ. The general agreement on the importance of detecting autocorrelation of regression errors has produced a large literature on the subject, and has led to the introduction of a variety of tests. The literature on testing for autocorrelation in linear regression models was surveyed in chapter 3, while chapter 4 outlined the limited progress which has been made towards documenting the consequences of model mis-specification for the properties of these tests.

In chapter 4 we emphasised the importance for applied workers of using tests which maintain their power when their underlying assumptions are violated in some commonly occurring ways. Such problems as departures from normality, omitted regressors or superfluous regressors should not seriously weaken a test for autocorrelation (or any test) if it is to be truly useful, as these violations are likely to occur and will typically be unknown to the applied researcher. This chapter explores one such scenario. Exact techniques are used to study the power functions of five tests for autocorrelation when they are applied to a model in which the disturbance variance is heteroscedastic in one of three different ways.

This work extends those of Epps and Epps (1977) and Giles and Small (1991) (both of which studied the power of the Durbin Watson test when the errors are heteroscedastic) to include the Berenblut and Webb (1973) test, the alternative DW test (King (1981)) and two versions of King's (1985) point optimal test[1].

The next section briefly sets out the models and tests used. Some theoretical results arising from these models are presented in section 6.3, which is followed by a description of the data used in the numerical evaluations. section 6.5 reports the main findings and section 6.6 offers some concluding comments.

6.2 Model and Tests

Consider the standard linear regression model

$$y = X\beta + u \tag{1}$$

where y is a (Tx1) vector of observations on the dependent variable, X is a (Txk) full rank non-stochastic regressor matrix, β is a (kx1) parameter vector and u is a (Tx1) vector of

[1] It will be recalled from chapter 3 that this point optimal test exploits results from Kadiyala (1970).

(possibly heteroscedastic) disturbances following the AR(1) process:

(2) $$u_t = \rho u_{t-1} + \varepsilon_t$$

where $|\rho| < 1$, $t = 1,\ldots,T$, and $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$ and serially independent. When $u=(u_1,\ldots,u_T)'$ is homoscedastic it has covariance matrix $E(uu')=\left(\frac{\sigma_\varepsilon^2}{1-\rho^2}\right)V$, where

$$V = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & & & \rho^{T-1} \\ \rho & 1 & & & & & \vdots \\ \rho^2 & \rho & \ddots & & & & \vdots \\ \vdots & & & \ddots & & & \vdots \\ \vdots & & & & \ddots & & \vdots \\ \vdots & & & & & \ddots & \vdots \\ \rho^{T-2} & & & & & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \cdots & & & \rho & 1 \end{bmatrix}.$$

We consider tests of $H_0 : \rho = 0$ against $H_a^+ : \rho > 0$ and $H_a^- : \rho < 0$ individually, each conducted at the 5% nominal size.

The statistics for each of the tests considered can be written as a ratio of quadratic forms in u, the general form of this ratio being

(3) $$r = \frac{u'Q\ u}{u'M\ u}$$

where $M = I_T - X(X'X)^{-1}X'$ and Q is some other non-stochastic (TxT) matrix, the form of which determines the individual test statistic.

(i) The Durbin Watson (DW) Test

This has

$$Q = MAM, \text{ where } A = \begin{bmatrix} 1 & -1 & 0 & . & . & . & 0 \\ -1 & 2 & -1 & & & & . \\ 0 & -1 & 2 & & & & . \\ . & & & . & & & . \\ . & & & & . & & 0 \\ . & & & & & 2 & -1 \\ 0 & 0 & . & . & 0 & -1 & 1 \end{bmatrix}.$$

The DW test is an approximately Uniformly Most Powerful (UMP) test of H_0 against H_a^+ for all design matrices, the approximation being due to a small modification made to the density function of the stationary AR(1) error process by Durbin and Watson (1950)[2]. In addition, when the columns of X are linear combinations of k of the eigenvectors of A, the DW test is an approximately Uniformly Most Powerful Invariant (UMPI) test[3] against H_a^+. Throughout this chapter invariance is with respect to an affine transformation of the dependent variable so that the DW test under this eigenvector condition is approximatley UMP among all those tests which use statistics which are invariant to transformations of the form $y^* = y\gamma_1 + X\gamma_k$, where γ_1 is a positive scalar and γ_k is a k×1 vector.

---

2 Durbin and Watson ignored the term $\alpha \exp\left[\frac{1}{2\sigma^2}\{\rho(1-\rho)u'C_1u\}\right]$ which occurs in the joint density function of u assuming (1) and (2). Here C_1 is a matrix of zeros apart from ones in the top left and bottom right corners, and α is a positive scalar.

3 Recall that Cassing and White (1983) provide a methodology for evaluating the validity of this eigenvector condition.

(ii) The Berenblut and Webb (BW) Test.

Berenblut and Webb (1973) proposed two tests, the first of which used Q = MBM, where

$$B = \begin{bmatrix} 2 & -1 & 0 & . & . & . & 0 \\ -1 & 2 & -1 & & & & . \\ 0 & -1 & 2 & & & & . \\ . & & & . & & & . \\ . & & & & . & & 0 \\ . & & & & & 2 & -1 \\ 0 & 0 & . & . & 0 & -1 & 1 \end{bmatrix} .$$

This statistic arose originally from a consideration of a modified form of the density function of non-stationary first order autoregressive errors. The second test statistic was arrived at by considering the likelihood ratio test of H_0 against $\bar{H}_a:\rho \neq 0$ (again in the context of a non-stationary AR(1) process) and replacing the inverse of the covariance matrix of the AR(1) process with B. This gave rise to a statistic using:

$$Q = B - BX(X'BX)^{-1}X'B \; .$$

This second test will be referred to as the BW test and was shown by Berenblut and Webb to possess the optimal power qualities of both the DW test and the first Berenblut and Webb test. In an empirical evaluation Berenblut and Webb found that the BW test was more powerful than the DW test at high values of ρ for six different design matrices.

(iii) The Alternative Durbin Watson (ADW) Test

Here

$$Q = MA_0M, \text{ where } A_0 = \begin{bmatrix} 2 & -1 & 0 & . & . & . & 0 \\ -1 & 2 & -1 & & & & . \\ 0 & -1 & 2 & & & & . \\ . & & & . & & & . \\ . & & & & . & & 0 \\ . & & & & & 2 & -1 \\ 0 & 0 & . & . & 0 & -1 & 2 \end{bmatrix}.$$

King (1981) proposed this test and found it to be a Locally Best Invariant (LBI) test in the neighbourhood of $\rho = 0$. In the same paper, King discussed results from an empirical comparison of the power functions of the DW and ADW tests, in which the latter generally performed better than the former against negative autocorrelation and for $\rho < 0.5$.

(iv) The Point Optimal ($s(\rho_1)$) Tests

This class of tests was introduced by King (1985) drawing on earlier work by Kadiyala (1970). For these tests the numerator matrix is given by

$$Q = \Sigma(\rho_1)^{-1} - \Sigma(\rho_1)^{-1}X\ \{X'\Sigma(\rho_1)^{-1}X\}^{-1}X'\Sigma(\rho_1)^{-1}$$

where $\Sigma(\rho_1)^{-1} = \dfrac{1}{1-\rho_1^2} V(\rho_1)$ and $V(\rho_1)$ is V with ρ fixed at some chosen value, ρ_1. King showed that this test is most powerful invariant (MPI) when the selected value for ρ_1 is the true value of ρ . He studied the power of two versions of the Point Optimal test, $\rho_1 = 0.5$ and $\rho_1 = 0.75$, and the DW, ADW and BW tests using a wide range of design matrices. King found that small power increases occurred when using either s(0.5) or s(0.75), in preference to DW or ADW, with large samples of smoothly evolving regressors; and more significant differences were apparent in smaller samples. In the case of Watson's (1955) matrix[4], the $s(\rho_1)$ and BW tests had power functions which approached unity as $\rho \to 1$, in contrast to the DW and ADW power functions which peaked at $\rho = 0.75$ and then declined.

6.3 Theoretical Discussion

The power function of each test against H_a^+ for a nominal significance level of $100\alpha\%$ and its associated critical value, $r^*(\alpha)$, can be found by found by substituting values of ρ into the expression

$$(4) \qquad \mathrm{pr}\Big\{ \, r < r^*(\alpha) \mid V = V(\rho) \, \Big\} \,.$$

The analogous expression for testing against H_a^- is $\mathrm{pr}\{r>r^*(\alpha)|V=V(\rho)\}$ but the appropriate critical values are different. The structure of each test statistic which is given by (3), allows the use of the well known manipulations, in the style of Koerts and Abrahamse (1969), for example, to write (4) as:

$$(5) \quad \mathrm{pr}\Big\{ \, r < r^*(\alpha) \mid V = V(\rho) \, \Big\} = \mathrm{pr}\Big\{ \, u'(Q-r^*M)u < 0 \mid V = V(\rho) \, \Big\}$$

$$= \mathrm{pr}\Big\{ \sum_{j=1}^{T} \lambda_j \, \chi_j^2 < 0 \Big\}$$

where the λ_j's are the eigenvalues of $(Q-r^*M)V$ and the χ_j^2's are independent central chi square variates, each with one degree of

[4] The columns of this matrix are given by a_1, $(a_2+a_T)/\sqrt{2}$, $(a_3+a_{T-1})/\sqrt{2}$, ..., where the a_i are the eigenvectors corresponding to the eigenvalues of the differencing matrix A arranged in increasing order.

freedom. The following result from Evans and King (1985a) provides a useful unification: if $Q = B - BX(X'BX)^{-1}X'B$ and $M=I-X(X'X)^{-1}X'$ then $QM = MQ$, so that $Q = MQM$. This allows us to write the BW and $s(\rho_1)$ tests as DW-type tests with a particular A matrix and so to represent the power of all of the tests considered here as depending on the eigenvalues of $M(A-r^*I)MV$ for some non-stochastic A. The form of (5) facilitates the computation of the power of each test for any covariance matrix using, for example, the FQUAD routine from Koerts and Abrahamse (1969) or Davies' (1980) algorithm[5]. The numerical evaluations reported below were conducted using a fortran version of Davies' algorithm contained in the SHAZAM (1993) computer package.

To allow for the simultaneous presence of heteroscedasticity and the AR process of (2), we draw on Theorem 5.1 of the previous chapter, which provides the justification for assuming white noise ε_t's and introducing heteroscedasticity into the model through the u_t's. The following forms of heteroscedasticity are considered

$$\text{(6)} \qquad \mathrm{var}(u_t) = cZ_t^{\alpha}$$

$$\text{(7)} \qquad \mathrm{var}(u_t) = c(1 + \gamma Z_t)$$

$$\text{(8)} \qquad \mathrm{var}(u_t) = \begin{cases} 1 & t \le T_1 \\ h & t > T_1 \end{cases} \qquad t = 1,2,\ldots,T_1,\ldots,T$$

where Z_t is a suitable transformation of the value of the jth

[5] The FQUAD routine is suitable only for problems such as (5). Davies' algorithm can be used more generally to find the cdf of a weighted sum of non-central chi-squares plus a standard normal variate.

regressor at time t, X_{jt}, and α, γ and h are selected constants. The proportionality constant $c = \frac{\sigma_\varepsilon^2}{1-\rho^2}$ does not influence the power or size of the tests considered and will not concern us further.

These three forms of heteroscedasticity are ones which are likely to occur in practice. Multiplicative heteroscedasticity, (6), has been considered by several writers, including Harvey (1976) who concentrated on the estimation of a model with this characteristic. The process is easily generalised to include dependence on more than one regressor but this study is restricted to the simple case outlined above. The additive heteroscedasticity of (7) can represent any situation where the variance of the dependent variable is assumed to be a linear function of some transformation of the regressors. A notable special case is the random coefficient model of Hildreth and Houck (1968). The third model of var(u) is likely to arise when the regression parameters exhibit a structural break and will be referred to as Chow heteroscedasticity after the well known F tests for this phenomenon, first discussed by Chow (1960).

Although the problems of autocorrelation and heteroscedasticity are usually treated separately, there is no good reason to assume homoscedasticity when a test for AR(1) errors is conducted. Heteroscedasticity of the forms (7) and (8), in particular, are likely to occur in time-series regressions, while testing for spatial autocorrelation when using cross-section data provides another motivation for this study.

We now turn to the appropriate form of the covariance matrix of u when both heteroscedasticity and autocorrelation are present. There are at least two possible structures for this matrix which might reasonably be considered. Giles and Small (1991) assume that the vector of disturbance variances is substituted into the leading diagonal of V, giving

$$V^* = \begin{bmatrix} \sigma_1^2 & \rho & \cdot & \cdot & \cdot & \cdot & \rho^{T-1} \\ \rho & \sigma_2^2 & & & & & \rho^{T-2} \\ \rho^2 & & & \cdot & & & \cdot \\ \cdot & & & & \cdot & & \cdot \\ \cdot & & & & & \cdot & \rho \\ \rho^{T-1} & \cdot & & \cdot & \cdot & \rho & \sigma_T^2 \end{bmatrix}.$$

This covariance matrix arises naturally, but not exclusively, from the Hildreth-Houck (1968) random coefficient model with AR(1) errors. It can also be derived from a model in which only the intercept is random and in this context some extra insight is gained. Suppose that

$$\begin{aligned} y_t &= X_t'\beta + \mu_t + u_t \\ &= X_t'\beta + \nu_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \end{aligned}$$

where $\mu_t \sim N(0,\sigma_\mu^2)$ is the random intercept and $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$ is independent of μ_t. Then $\nu_t = \mu_t + u_t$ has covariance matrix given by cV^* with $\sigma_t^2 = (1 + \frac{\sigma_\mu^2}{k})\ \forall\ t$. This error components model is a special case of models in which the dependent variable is observed only with errors (such as those considered by Pagan

(1973)). If we impose homoscedasticity by assuming that $\sigma_t^2 = \lambda$, then V* reflects the following autoregressive process

(9) $$\lambda u_t = \rho^s u_{t-s} + \varepsilon_t \qquad t = 1,\ldots,T$$

in which the first autocorrelation is ρ/λ, while all subsequent autocorrelations are ρ. Since $\lambda > 1$ the first autocorrelation is weaker than all the others and this reduces the average power of all tests considered here, irrespective of the data, as the following theorem shows.

Theorem 6.1

When the autoregressive process is given by (9) rather than (2) the average value of the test statistic for all tests considered here is increased when testing against H_a^+.

Proof

Let $S = M(A-r^*I)M$ and define the ij^{th} element of S by s_{ij}. Consider the first moment of $(r-r^*)$ which is given by $E(u'S\,u) = tr(SV)$. We can decompose V* as

$$V^* = V + \Lambda$$

where $\Lambda = diag(\lambda^*)$

and $\lambda^* = \lambda - 1 > 0$.

If V* is the true covariance matrix of u then we must compare tr(SV) with tr(SV*).

$$tr(SV^*) = tr(S(V+\Lambda))$$

$$= tr(SV) + \sum_{i=1}^{T} s_{ii}\lambda^*$$

$$= tr(SV) + \lambda^* \; tr(S) \; .$$

Observe that $tr(S) = E(r-r^*)|_{\rho=0}$ and recall that $E(r)|_{\rho=0} > r^*$ when testing against H_a^+. We can conclude that $tr(S) > 0$ and therefore that $tr(SV^*) > tr(SV)$. #

So, at least on average, the probability of rejecting the null hypothesis in favour of H_a^+ is reduced as λ increases. When the alternative hypothesis is H_a^-, the null is rejected for values of r which are smaller than r^*. Thus, $E(r)|_{\rho=0} < r^*$ and therefore $tr(S) < 0$, leading to the conclusion that the average value of the test statistic is reduced with the same power effects as those for H_a^+.

This theorem illuminates the discussion between Revanker(1980) and King (1982) referred to in the previous chapter. It is clear that the power of a DW test (and all of the other tests studied in this chapter) is lower in the error components model than in a similar model with a standard error formulation. This power reduction occurs in finite samples as well as asymptotically. There is, however, no size distortion induced so that the probability of a type I error is unchanged. The question of whether a DW type test is appropriate or not will be deferred to allow the numerical evaluations reported later in the chapter to be considered.

Because the scale of λ matters for the power of the test it was considered important in the numerical evaluations of test power under heteroscedasticity to standardise the scale of the σ_t^2 entering the leading diagonal of V^*. Transformations used in constructing Z_t were chosen such that the σ_t^2 were constrained to take a minimum value of unity, while still reflecting the variability of X_j. For (6) $Z_t = \dfrac{X_{jt}}{\min(X_{jt})}$, while (7) has $Z_t = \dfrac{X_{jt} - \min(X_{jt})}{\max(X_{jt}) - \min(X_{jt})}$.

As outlined in the previous chapter, the major alternative to joint modelling of heteroscedasticity and autocorrelation by means of V* is approached by using the elementary expressions for the correlations between two variables and imposing the AR(1) process. This results in the following covariance matrix, neglecting the scale factor:

$$V^{**} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 & \cdots & \rho^{T-1}\sigma_1\sigma_T \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & & \rho^{T-2}\sigma_2\sigma_T \\ \rho^2\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho^{T-1}\sigma_1\sigma_T & \cdot & \cdot & \cdots & \sigma_T^2 \end{bmatrix}.$$

Both V* and V** have been considered throughout the research reported in this chapter. It will be shown below that the form of the covariance matrix matters a great deal for the properties of the tests considered here.

The data dependence of the distribution of r precludes direct analytical evaluation of the powers of the tests under consideration in almost all cases. Some results are obtainable, however, at the boundaries of the parameter space by examining the limits of the eigenvalues of (5) as $\rho \to \pm 1$. This technique was used by Krämer (1985) to prove that, for regressions with no intercept, the limiting power of the DW test as $\rho \to 1$ is always zero or unity. A more involved, but similar, analysis enabled Zeisel (1989) to show that when an intercept is present the power of the DW test as $\rho \to 1$ lies strictly between these two values. More recently, Small (1993) has generalised both of these results to the ADW, BW and $s(\rho_1)$ tests.

This section presents two further limiting power results which apply when the disturbances are heteroscedastic in particular ways.

Theorem 6.2

When $\text{var}(u_t) = cZ_t^{\alpha}$ and cov(u) is given by V**, the limiting power of all tests considered here, as $\rho \to 1$, is zero or unity unless $Z^{\alpha/2}$ is in the column space of X, irrespective of the presence of an intercept.

Proof

Under the conditions of the theorem, the covariance matrix of u is given by

$$V^{**} = \begin{bmatrix} Z_1^{\alpha} & (Z_2Z_1)^{\alpha/2} & . \; . \; . \; . & (Z_TZ_1)^{\alpha/2} \\ (Z_1Z_2)^{\alpha/2} & Z_2^{\alpha} & . & (Z_TZ_2)^{\alpha/2} \\ (Z_1Z_3)^{\alpha/2} & (Z_2Z_3)^{\alpha/2} & Z_3^{\alpha} & \vdots \\ \vdots & & \ddots & \\ (Z_1Z_T)^{\alpha/2} & . \; . \; . \; . \; . \; . \; . & . \; . \; . \; . & Z_T^{\alpha} \end{bmatrix}$$

$$= \begin{bmatrix} X_{j1}^{\alpha/2} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ X_{jT}^{\alpha/2} \end{bmatrix} \begin{bmatrix} X_{j1}^{\alpha/2} & . & . & . & . & X_{jT}^{\alpha/2} \end{bmatrix}.$$

Recall that $X_j^{\alpha/2}$ is not in the column space of X by assumption. This implies that, in general, MV** $\neq$ 0. Notice that V** has rank equal to unity so that M(A-r*I)MV** has only one non-zero eigenvalue. The sign of this eigenvalue uniquely determines the power of the test, a positive eigenvalue implying a limiting power of zero. #

This result is extended in two directions by the following corollaries.

Corollary 1

The limiting powers, as $\rho \to 1$ of all tests considered, when var(u_t) are given by (7) or (8) and cov(u) = V**, is also zero or unity.

Proof

When var(u_t) is given by (7) then, as $\rho \to 1$, V** $\to \nu\nu'$, where $\nu = \left(\sqrt{1+\gamma Z_1}\ \sqrt{1+\gamma Z_2}\ .\ .\ .\ .\sqrt{1+\gamma Z_T} \right)'$ is not spanned by the columns of X in general.

When var(u_t) is given by (8) and $\rho \to 1$, V** $\to$ ww' where w=(1,1,...,1,h,h,...,h)' is also not spanned by the columns of X. In both of these cases V** also has unit rank so that the limiting powers of all tests are either zero or unity. #

Corollary 2

When cov(u) = V** and var(u_t) is given by either (6), (7) or (8), the limiting powers of all tests considered are always either zero or unity as $\rho \to -1$.

Proof

If $V^{**} \to \nu\nu'$ with $\nu = (\nu_1, \nu_2, \ldots, \nu_T)'$ then as $\rho \to -1$, $V^{**} \to \nu^-\nu^{-\prime}$ with $\nu^- = (\nu_1, -\nu_2, \nu_3, -\nu_4, \ldots, (-1)^{T-1}\nu_T)'$. Now if ν is not in the column space of X, then neither is ν^- but V** has rank approaching unity as $\rho \to -1$ under these conditions. The power must therefore approach zero or unity as $\rho \to -1$. #

To conclude this section we establish the following theorem which concerns the limiting power when covariance matrix V* is appropriate.

Theorem 6.3

When var(u_t) = $k(1+\gamma Z_t)$ and cov(u) is given by V*, the limiting powers of all of the tests considered here as $\rho \to 1$ are constant for all $\gamma > 0$, providing an intercept is included in the regression model.

Proof

Under the conditions of Theorem 6.3 we can decompose the covariance matrix of u as

$$V^* = \Sigma + \gamma\Lambda$$

where $\Sigma = ii'$ for $i = (1,1,\ldots,1)'$, $\Lambda = \text{diag}(Z_t)$ and γ is a scalar. Let $S = M(A - r^*I)M$ and consider the eigenvalues of SV^*, being the Λ which satisfy

$$\lambda w = SV^*w\ , \qquad \text{for some non-null vector } w,$$

or:

$$\lambda w = S(\Sigma + \gamma\Lambda)w.$$

Now when an intercept is present, $M\Sigma = S\Sigma = 0$ so that

$$\lambda w = \gamma S\Lambda w.$$

Consider some other non-zero scalar, γ^*. We can write

$$\lambda w = \gamma^* \left[\frac{\gamma}{\gamma^*}\right] S\Lambda w\ ,$$

so that

$$\frac{\gamma}{\gamma^*}\lambda w = \gamma^* S\Lambda w.$$

Thus, altering the value of γ scales each eigenvalue by the same factor, which does not affect the rejection probability.

6.4 The Data

Several regressor matrices were used in an empirical study in an attempt to reveal the effects of data characteristics on the tests' powers under the mis-specifications outlined above. The design matrices were of two sizes, 60x3 and 20x3, and are characterised as follows:

X1: A constant and the income and price series from Durbin and Watson's (1951) consumption of spirits example.

X2: A constant, the quarterly Australian Consumer's Price Index commencing 1959(1), and the same series lagged one period.

X3: A constant, a linear time trend and observations drawn from the Normal(30,4) distribution.

X4: A constant, a linear time trend and observations drawn from the Uniform[0,10] distribution.

X5: A constant, a linear time trend and observations drawn from the Lognormal (2.226, 19.58) distribution.

X6: a_1, $(a_2 + a_T)/\sqrt{2}$, $(a_3 + a_{T-1})/\sqrt{2}$, where a_1 ,..., a_T are the eigenvectors corresponding to the eigenvalues of A arranged in ascending order. Note that a_1 is a constant as it corresponds to the zero root of A.

These design matrices were chosen to represent a range of data characteristics. The slowly evolving X1 matrix involves annual data while X2 has a weak seasonal pattern. Both X1 and X2 have been used in previous studies in the general field of autocorrelation testing (*e.g.*, King (1985) and Evans (1992)) Several previous studies have suggested the use of artificial data of various types. The lognormal data, for example, are often used to represent cross section data which are known to be skewed and therefore particularly relevant to scenarios involving heteroscedasticity. The X6 matrix was shown by Watson (1955) to produce the most inefficient OLS estimates within the class of orthogonal design matrices.

For each X matrix one regressor was selected to be the X_j of (6) and (7).The variables used for this purpose were: Income (XI), CPI (X2), Normal (X3), Uniform (X4), Lognormal (X5) and $(a_2+a_T)/\sqrt{2}$ (X6). The scalars α and γ were chosen to give desired values of the ratio

$$h = \frac{\text{maximum var}(u_t)}{\text{minimum var}(u_t)} .$$

The degree of heteroscedasticity introduced was controlled in this way with h being set at six values ranging from 1.0 (which implies homoscedasticity) to 2.5. In the other exact work of this type, Epps and Epps (1977) and Giles and Small (1991) used the same measurement criterion for heteroscedasticity, although other measures could be considered, such as the coefficient of variation.

No size corrections were made to the heteroscedastic power functions. The reason for this is that the purpose of this study is to determine the effect of heteroscedastic disturbances on the power of each test when based on least squares residuals. If an applied worker knew that the disturbances were heteroscedastic she would use a GLS type estimator which would account for the heteroscedasticity and simultaneously return the nominal autocorrelation test to its true size. This study presumes that the researcher is ignorant of the complications due to heteroscedasticity.

6.5 Numerical Results

For convenience, we discuss the results of the numerical study in two groups, distinguished by the covariance matrix used.

6.5.1 Results Using V*

This section discusses the results obtained by using the first version of the mis-specified covariance matrix, V*, introduced above. For all of the Figures and Tables referred to the sample size is 20 observations unless otherwise specified; this additional material is located at the end of this chapter.

The correctly specified (h=1.0) power functions (*e.g.*, Figures 6.1a and 6.1b) are in accord with the findings of previous studies. In the case of X6 and T = 60 the results are identical to those reported by King (1985). As expected, the degree of extra power available from selecting the best test, rather than the worst, varies considerably with the data used. This can be seen by comparing Figure 6.1a, where there is very little difference between the tests, with Figure 6.1b, which is based on different data and shows that large power gains are available by selecting the best test. In a correctly specified model when using any data from X1 to X4 inclusive and any given AR parameter, there are only minor power advantages obtainable by choosing the best test from the set studied. This can be seen from the h=1.0 columns of Tables 6.1 to 6.4. The correctly specified power functions for X1, X2 and X3 are almost identical to those of X4 which appear in Figure 6.1a. This graph also confirms that the ADW test is relatively weak for S(0.75) and BW tests have very similar power functions which, as a group, dominate the DW and ADW tests over this region. These rankings are reversed, however, for tests against H_a^-.

Greater differences between the tests are evident when using X5 and X6 in a correctly specified model. Table 6.5 is based on a sample size of 20 and shows that for X5 with ρ=0.8 and h=1.0, the BW test has a power of 0.816, which is 5% higher than the 0.773 power of the weakest test (ADW). The same comparison using X6 reveals a massive 84% increase in power from using the BW test rather than the considerably weaker DW test (see Table 6.6). Figure 6.1b clearly illustrates the extreme differences which can arise even in correctly specified models. As was noted by King (1985), when using Watson's matrix, the power functions of the DW and ADW tests peak at ρ=0.7 and then decline as $\rho \to 1$. This is in stark contrast to the $s(\rho_1)$ and BW tests whose powers are strictly increasing in ρ, for this matrix, and is likely to be due to the effect on the residuals of very inefficient OLS parameter estimates.

When heteroscedasticity is introduced there is potential for the true size of each test to be distorted away from the nominal (here 5%) level. It is obviously more difficult to compare the power functions of two tests (or the same test in two different models) when one has a higher Type I error probability than the other. The degree of size distortion encountered in this study varied with h, the true sizes of all tests in all models falling within the following ranges: [0.049, 0.052] when h ≤ 1.2, [0.049, 0.055] for h=1.5 and [0.048, 0.058] for h = 2.0. These distortions are small, relative to the power changes that are induced by increasing h above unity, which allows valid comparisons of power functions across tests and across different degrees of heteroscedasticity.

The effect of introducing a moderate (h=1.5) degree of heteroscedasticity can be seen in Tables 6.1 to 6.6, in which the sample size is always 20. Figures 6.3 and 6.4 present this information, and power functions for other values of h, graphically. These figures also include powers for models with different forms of heteroscedasticity, although the scale of the variances has been standardised.

The most notable feature of Figures 6.3 and 6.4 is that while all tests have homoscedastic power functions which are strictly increasing in ρ, for these design matrices, all these power functions are declining in ρ as $\rho \to 1$ for all $h > 1$ when $T = 20$. Notice also that the decline in power as h increases is consistent with the effect predicted by Theorem 6.1.

Figure 6.4b graphs the power of the s(0.5) test for various degrees of additive heteroscedasticity with $T = 20$ and using X4. The effect predicted by Theorem 6.3 is clearly evident in this diagram. This data set illustrates the least severe power reductions encountered as a consequence of heteroscedasticity. Again, the power differences between the tests for a given h are relatively small, as can be seen from Table 6.4.

The effect that sample size has on the power of these tests was noted by King (1985). Other things constant, a larger sample size increases the power of each test and reduces the power differences between the tests. In addition, Figures 6.3a and 6.3b show that the DW test is much more robust to small degrees of heteroscedasticity when $T = 60$ than when $T = 20$. This effect was found to be common to all tests and all design matrices.

It can be seen from Figures 6.3 and 6.4 that the precise form of $var(u_t)$ is not important for the general shape of the mis-specified power functions. The crucial determinant of the serious power losses evident in these graphs is the scale of the leading diagonal elements of V*, as suggested by Theorem 6.1. Further weight is given to this conclusion by Figure 6.4a, which plots power curves for the DW test using X1 with T = 20. In this figure, n represents the number of non-unity leading diagonal elements of V*, all such elements taking a value of 2.5. The conclusion is that as the average of the diagonal elements increases, the power of the test falls.

The ranking of the tests on the basis of their powers can change as a result of increasing h above unity. For example Table 6.5 shows that for X5 and h = 1.0 the limiting power of the BW test is superior to that of the ADW test while when h = 1.5 these rankings are reversed (recall that BW is LMPI as $\rho \to 1$). Comparisons such as this are potentially dangerous however, as they divert attention from the major effect of this form of heteroscedasticity and can give false confidence in the strength of one particular test.

6.5.2 Results using V**

In this section we consider the power functions when the true covariance matrix is V**. For these models, Tables 6.7 and 6.8 and Figures 5 and 6 provide selected graphs and tabulated power values, all of which are based on a sample size of 20. In Tables 6.8 and 6.7 heteroscedasticity is of the form given by (6) with the same number of observations (10) in each variance regime.

When the covariances between individual disturbances reflect their heteroscedasticity the true covariance matrix is given by V**. The scale of the disturbance variances is irrelevant in this case, as it affects all elements of the covariance matrix equally and can therefore be factored out. Another way of looking at this is to note that there is no implied mis-specification of the AR(1) process, in contrast to the V* case considered above.

The effect of this type of mis-specification on the sizes of the tests considered is minimal. The true sizes of all tests over almost all models were found to be identical to their nominal sizes to two decimal places. The exceptions to this were minor, with true sizes falling in the range (0.056, 0.060) for high values of the heteroscedasticity parameter when the data matrix was X2 or X6 (the nominal size was 0.050).

Disregarding, for now, the extremes of the parameter space, the powers of the tests were generally not significantly altered by introducing heteroscedasticity of the form V**. In cases where, for given ρ, the power of a test changed by more than ± 0.01, it was found that the data matrix used was the important factor, rather than the particular test.

These slightly larger deviations from correctly specified power functions occurred with X1 (for mid-range $\rho > 0$), X3 (strongly negative ρ), X4 (as $\rho \to 1$), X5 (as $\rho \to 1$) and X6 (mid-range and strongly negative ρ and as $\rho \to 1$). The most serious loss of power occurred with X6 for $0.055 < \rho < 1$.

It is interesting to note that although the BW, s(0.5) and s(0.75) tests are not intended for use against H_a^-, they have higher power against this alternative than against H_a^+ for given absolute values of ρ. As usual there is an exception to this statement which is provided by Watson's data set, X6.

The Tables show power values for $\rho=\pm1$. These were calculated using the techniques suggested by Krämer and Zeisel (1990). As shown in Theorem 6.2, for all tests considered here, when the true covariance matrix is V** and var(u) is given by (6), (7) or (8) the limiting power (in either direction) is always zero or unity. This fact accounts for significant deviations from correctly specified power functions as $\rho \to 1$ for X3, X4, X5 and X6, since the limiting V** power for all tests using these data sets is zero. Again, the precise form of var(u_t) is less important than the structure of the covariance matrix.

The real data produced limiting powers of unity, which provokes speculation as to the reasons for the difference from the artificial data in this respect.

To summarise this section, it has been found that mis-specification of the type given by V** has very little effect on the size and power of all tests studied unless the AR(1) parameter is very large in absolute value. In particular, the power of all tests approached zero as $\rho \to 1$ when artificial data were used.

6.6 Conclusion

This chapter has considered the effect that heteroscedastic errors have on the power of some tests for AR(1) errors, and has found severe reductions in power under two different covariance matrix structures and three types of heteroscedasticity.

When the underlying AR(1) process is altered by the introduction of heteroscedasticity of the V* type, the power of each test is lower (for all ρ) the greater is the scale of the disturbances, irrespective of the particular scheme for $var(u_t)$.

The second heteroscedastic covariance matrix used, V**, allows the AR(1) process to dominate, with the covariance terms reflecting the heteroscedasticity. In this case the most notable effect is on the limiting power as $\rho \to \pm 1$. Independently of the presence of an intercept, the limiting power of each test considered is either zero or one under these conditions, when various forms of heteroscedasticity contaminate the error process. The clear conclusion arising from this chapter is that there is no guarantee that the popular tests for AR(1) disturbances studied have any significant power when there is heteroscedasticity present. Furthermore, in many such cases the probability of detecting autocorrelation declines as ρ increases, and the consequences of ignoring it get more severe.

TABLE 6.1 Power of DW and s(0.5) Tests; Spirits Data (X1) Multiplicative Heteroscedasticity with V*				
	DW		s(0.5)	
ρ	h=1.0	h=1.5	h=1.0	h=1.5
-1.0	1.000	0.102	1.000	0.765
-0.8	0.922	0.085	0.932	0.086
-0.6	0.737	0.073	0.749	0.073
-0.4	0.434	0.063	0.441	0.064
-0.2	0.174	0.056	0.175	0.056
0.0[1]	0.050	0.050	0.050	0.051
0.0	0.050	0.049	0.050	0.049
0.2	0.174	0.054	0.176	0.054
0.4	0.405	0.059	0.409	0.060
0.6	0.653	0.063	0.656	0.063
0.8	0.818	0.062	0.819	0.062
1.0	0.990	0.049	0.884	0.059

1. The figures in this row show the true size of the test against a negative alternative. The following row is for tests against H_a^+.

TABLE 6.2
Power of ADW and BW Tests; CPI Data (X2)
Multiplicative Heteroscedasticity with V*

	ADW		BW	
ρ	h=1.0	h=1.5	h=1.0	h=1.5
-1.0	1.000	0.698	1.000	0.659
-0.8	0.935	0.768	0.898	0.720
-0.6	0.756	0.582	0.710	0.542
-0.4	0.448	0.336	0.422	0.317
-0.2	0.177	0.145	0.172	0.141
0.0[1]	0.050	0.051	0.050	0.051
0.0	0.050	0.049	0.050	0.050
0.2	0.173	0.138	0.170	0.136
0.4	0.398	0.285	0.397	0.285
0.6	0.635	0.436	0.645	0.443
0.8	0.793	0.488	0.809	0.502
1.0	0.859	0.066	0.876	0.071

1. The figures in this row show the true size of the test against a negative alternative. The following row is for tests against H_a^+.

TABLE 6.3 Power of DW and s(0.75) Tests; Normal Data (X3) Additive Heteroscedasticity with V*				
	DW		s(0.75)	
ρ	h=1	h=1.5	h=1	h=1.5
-1.0	1.000	0.686	1.000	0.491
-0.8	0.921	0.744	0.831	0.570
-0.6	0.737	0.561	0.646	0.431
-0.4	0.437	0.324	0.391	0.260
-0.2	0.175	0.141	0.166	0.123
0.0[1]	0.050	0.050	0.050	0.049
0.0	0.050	0.053	0.050	0.051
0.2	0.171	0.142	0.171	0.131
0.4	0.396	0.293	0.403	0.266
0.6	0.638	0.449	0.658	0.412
0.8	0.799	0.510	0.826	0.461
1.0	0.866	0.085	0.885	0.054

1. The figures in this row show the true size of the test against a negative alternative. The following row is for tests against H_a^+.

TABLE 6.4 Power of DW and s(0.75) Tests; Uniform Data (X4) Additive Heteroscedasticity with V*				
	DW		s(0.75)	
ρ	h=1.0	h=1.5	h=1.0	h=1.5
-1.0	1.000	0.826	1.000	0.818
-0.8	0.896	0.841	0.884	0.829
-0.6	0.707	0.650	0.695	0.638
-0.4	0.421	0.382	0.414	0.376
-0.2	0.172	0.161	0.171	0.159
0.0[1]	0.050	0.051	0.050	0.051
0.0	0.050	0.050	0.050	0.050
0.2	0.165	0.151	0.165	0.153
0.4	0.381	0.337	0.389	0.346
0.6	0.626	0.545	0.644	0.564
0.8	0.798	0.668	0.817	0.692
1.0	0.865	0.044	0.880	0.054

1. The figures in this row show the true size of the test against a negative alternative. The following row is for tests against H_a^+.

TABLE 6.5[1] Power of ADW and BW tests; Lognormal Data (X5) Chow-type Heteroscedasticity with V*				
	ADW		BW	
ρ	h=1.0	h=1.5	h=1.0	h=1.5
-1.0	1.000	0.721	1.000	0.706
-0.8	0.921	0.766	0.879	0.726
-0.6	0.736	0.577	0.689	0.543
-0.4	0.436	0.335	0.411	0.319
-0.2	0.175	0.147	0.170	0.144
0.0[2]	0.050	0.053	0.050	0.053
0.0	0.050	0.051	0.050	0.048
0.2	0.168	0.135	0.163	0.128
0.4	0.380	0.272	0.386	0.270
0.6	0.611	0.411	0.642	0.429
0.8	0.773	0.459	0.816	0.495
1.0	0.837	0.090	0.880	0.081

1. Ten observations were included in each of the two variance regimes.
2. The figures in this row show the true size of the test against a negative alternative. The following row is for tests against H_a^+.

TABLE 6.6[1] **Power of ADW and BW tests; Watson's Data (X6)** **Chow-type Heteroscedasticity with V***				
	ADW		BW	
ρ	h=1.0	h=1.5	h=1.0	h=1.5
-1.0	0.000	0.067	0.000	0.032
-0.8	0.477	0.328	0.253	0.190
-0.6	0.462	0.339	0.303	0.236
-0.4	0.311	0.238	0.242	0.193
-0.2	0.147	0.124	0.133	0.114
0.0[2]	0.050	0.051	0.050	0.051
0.0	0.050	0.051	0.050	0.051
0.2	0.146	0.123	0.129	0.111
0.4	0.299	0.227	0.317	0.234
0.6	0.434	0.308	0.595	0.409
0.8	0.450	0.278	0.828	0.536
1.0	0.301	0.097	0.936	0.099

1. Ten observations were included in each of the two variance regimes.
2. The figures in this row show the true size of the test against a negative alternative. The following row is for tests against H_a^+.

TABLE 6.7
Selected Powers - Multiplicative Heteroscedasticity with V**

	DW		s(0.5)		BW	
ρ	h=1.0	h=1.5	h=1.0	h=1.5	h=1.0	h=1.5
Spirits Data (X1)						
0.0	0.050	0.049	0.050	0.049	0.050	0.049
0.2	0.174	0.171	0.174	0.171	0.171	0.167
0.4	0.405	0.400	0.405	0.400	0.398	0.391
0.6	0.653	0.647	0.653	0.647	0.648	0.641
0.8	0.818	0.813	0.818	0.813	0.819	0.813
1.0	0.883	1.000	0.883	1.000	0.892	1.000
CPI Data (X2)						
0.0	0.050	0.052	0.050	0.050	0.050	0.050
0.2	0.171	0.176	0.173	0.172	0.170	0.170
0.4	0.396	0.401	0.402	0.400	0.397	0.395
0.6	0.638	0.642	0.647	0.644	0.645	0.643
0.8	0.799	0.802	0.808	0.806	0.809	0.808
1.0	0.866	1.000	0.873	1.000	0.876	1.000
Normal Data (X3)						
0.0	0.050	0.051	0.050	0.051	0.050	0.052
0.2	0.170	0.173	0.172	0.174	0.169	0.173
0.4	0.399	0.401	0.405	0.407	0.400	0.404
0.6	0.651	0.653	0.659	0.659	0.656	0.659
0.8	0.819	0.820	0.824	0.824	0.826	0.827
1.0	0.881	0.000	0.883	0.000	0.886	0.000

TABLE 6.8
Selected Powers - Multiplicative Heteroscedasticity with V**

	DW		s(0.5)		BW	
ρ	h=1	h=1.5	h=1	h=1.5	h=1	h=1.5
Uniform Data (X4)						
0.0	0.050	0.050	0.050	0.050	0.050	0.050
0.2	0.170	0.170	0.173	0.173	0.170	0.170
0.4	0.397	0.396	0.405	0.405	0.400	0.399
0.6	0.649	0.647	0.658	0.657	0.656	0.654
0.8	0.819	0.816	0.824	0.822	0.826	0.823
1.0	0.881	0.000	0.884	0.000	0.887	0.000
Lognormal Data (X5)						
0.0	0.050	0.049	0.050	0.050	0.050	0.050
0.2	0.165	0.163	0.167	0.167	0.163	0.162
0.4	0.381	0.378	0.392	0.391	0.386	0.383
0.6	0.626	0.622	0.644	0.643	0.642	0.639
0.8	0.798	0.795	0.814	0.813	0.816	0.815
1.0	0.865	0.000	0.877	0.000	0.880	0.000
Watson's Data (X6)						
0.0	0.050	0.050	0.050	0.051	0.050	0.051
0.2	0.142	0.140	0.142	0.143	0.129	0.130
0.4	0.291	0.287	0.339	0.339	0.317	0.317
0.6	0.424	0.418	0.604	0.603	0.595	0.594
0.8	0.440	0.434	0.819	0.817	0.828	0.826
1.0	0.308	0.000	0.925	1.000	0.936	1.000

Figure 6.1a

Uniform Data; T=20

h = 1.0

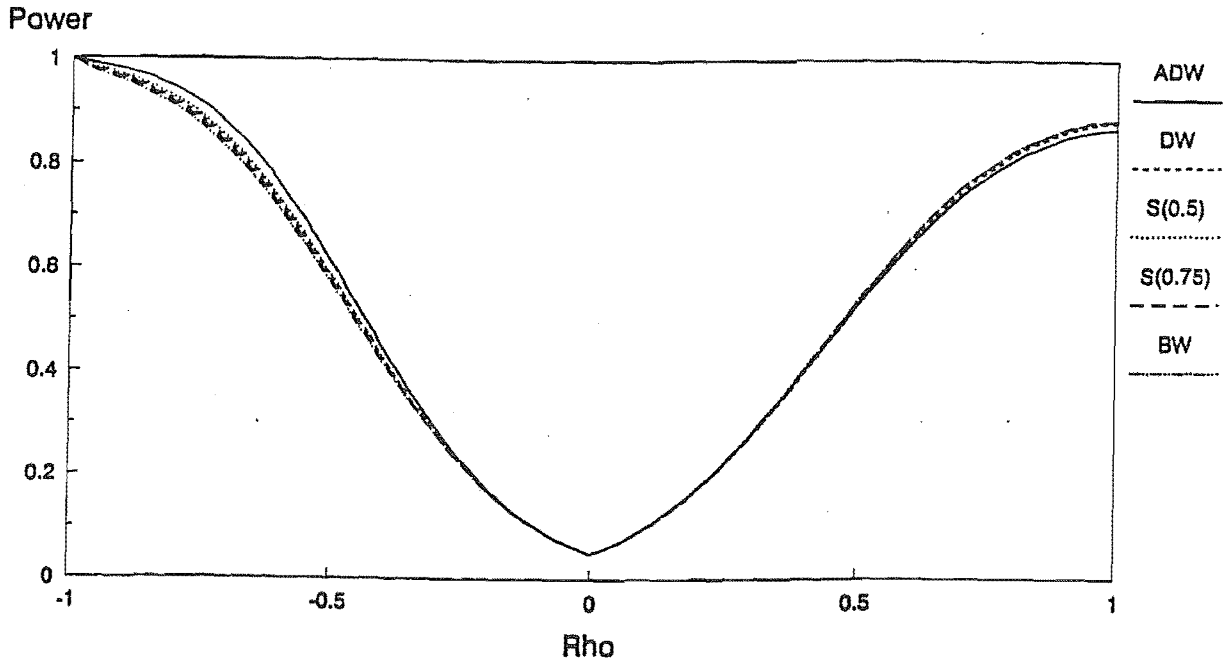


Figure 6.1b

Watson's Data; T=20

h = 1.0

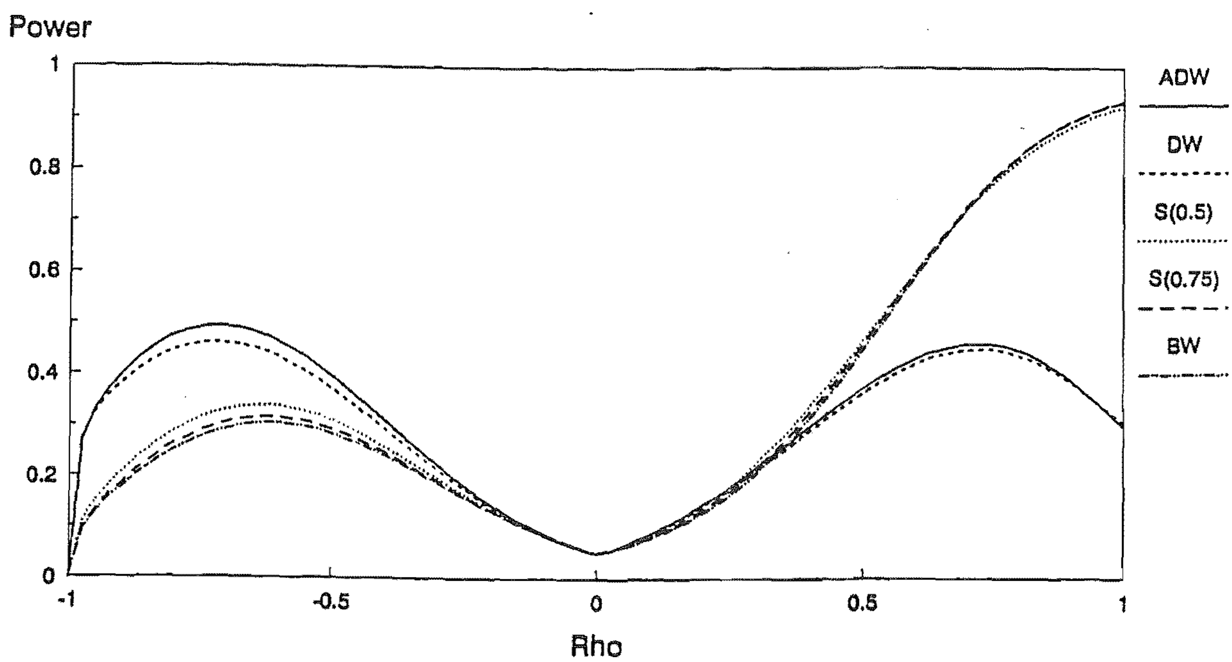


Figure 6.2a
Uniform Data; T=20
h = 1.5

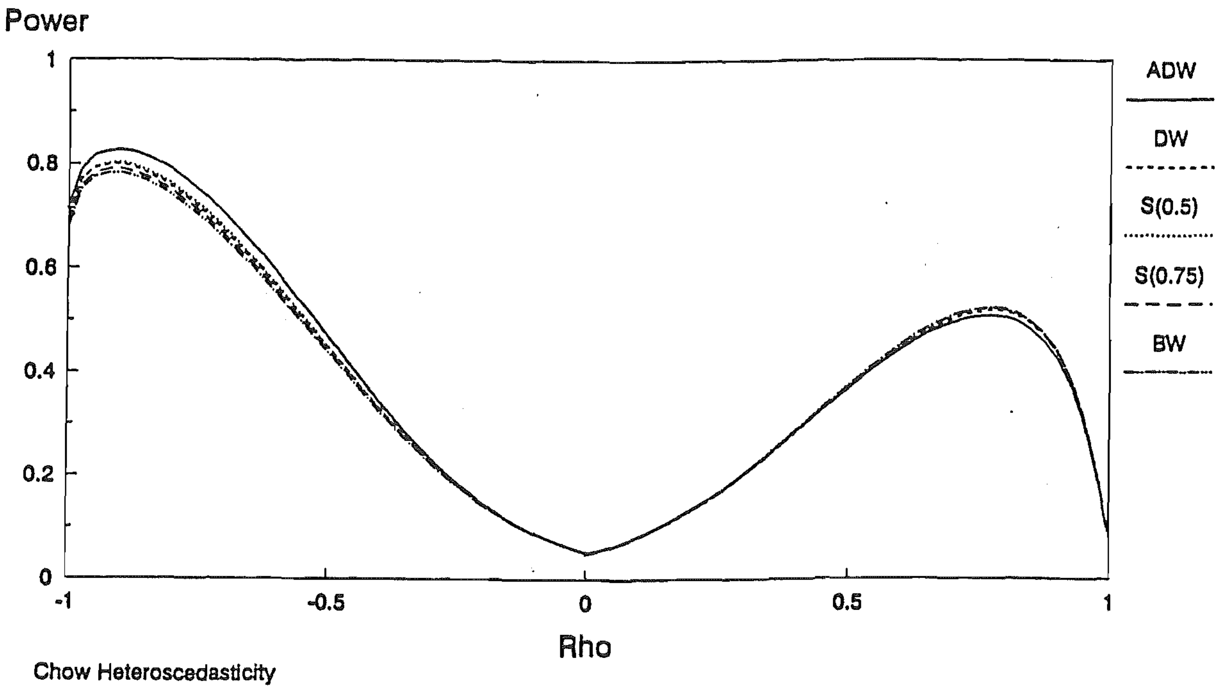


Figure 6.2b
Watson's Data; T=20
h = 1.5

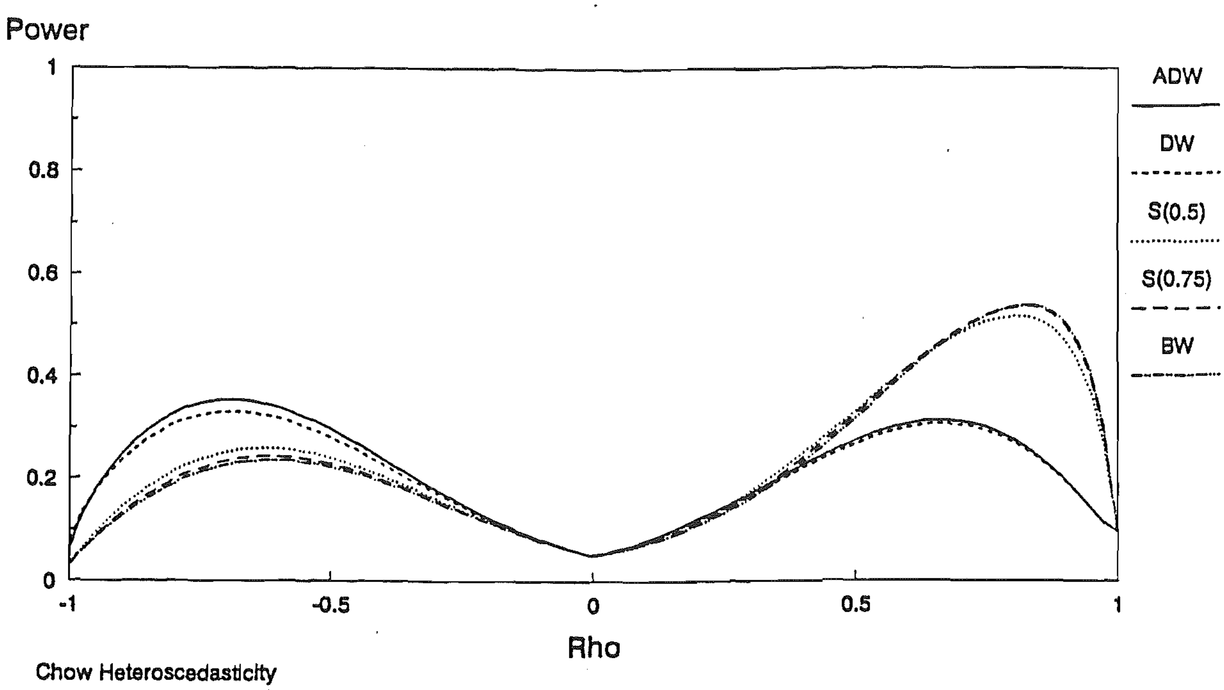
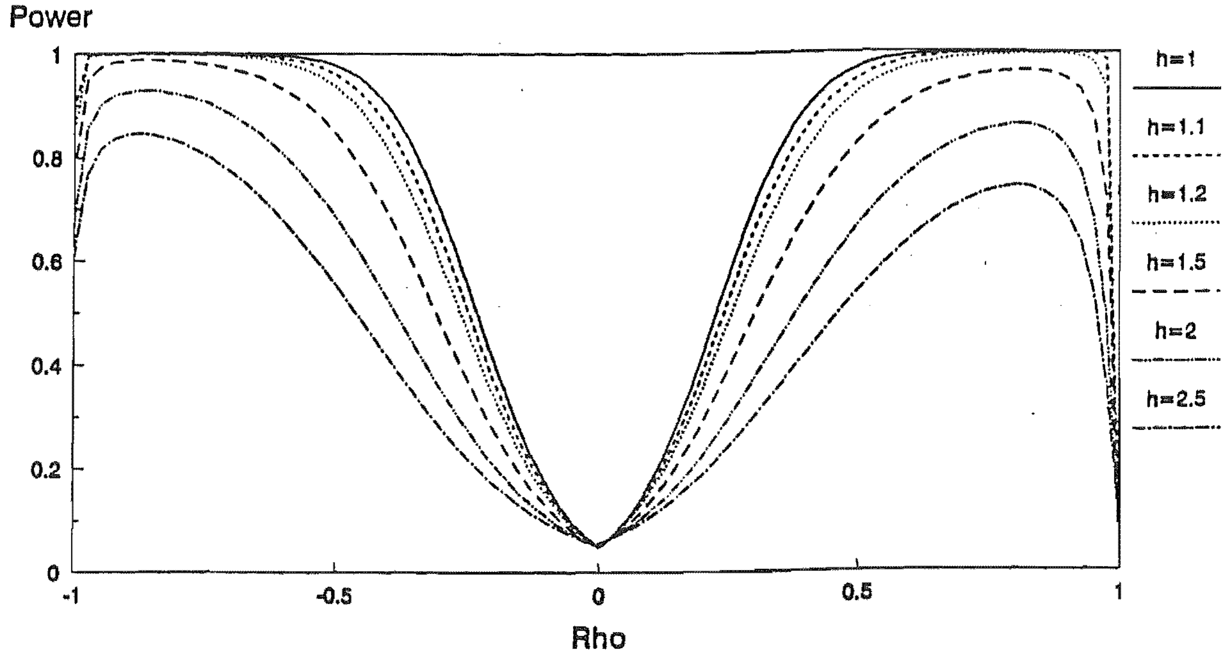
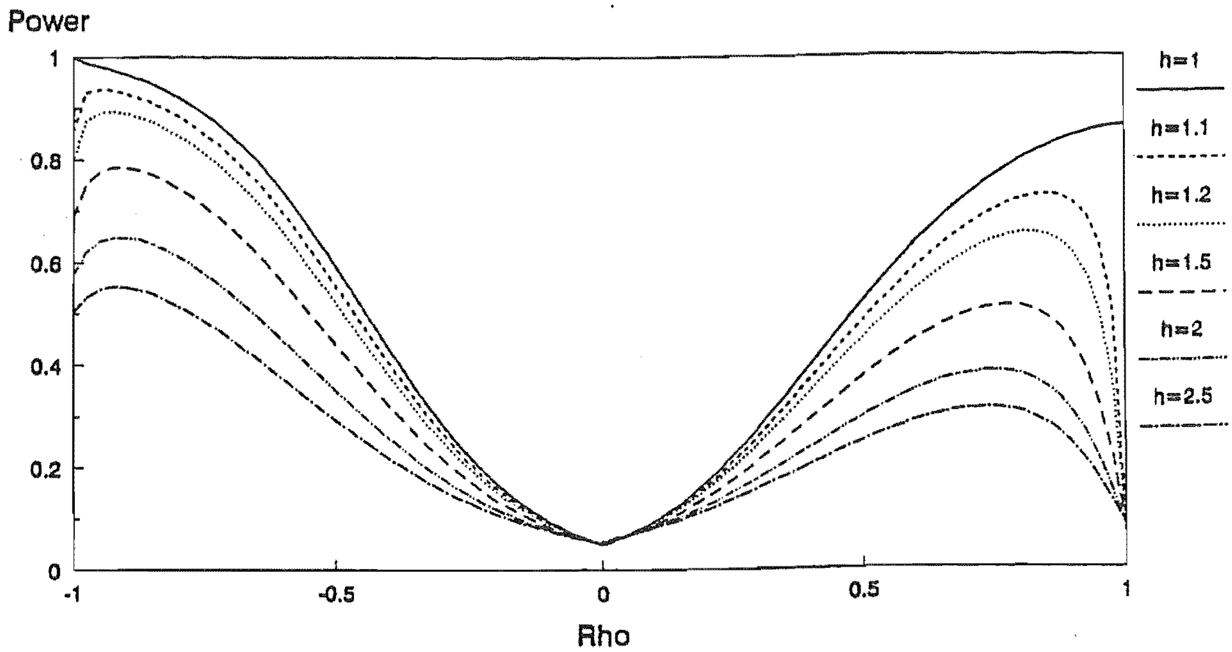


Figure 6.3a
CPI Data; T = 60
DW Test



Additive Heteroscedasticity

Figure 6.3b
CPI Data; T = 20
DW Test



Additive Heteroscedasticity

Figure 6.4a
Spirits Data; T = 20
DW Test

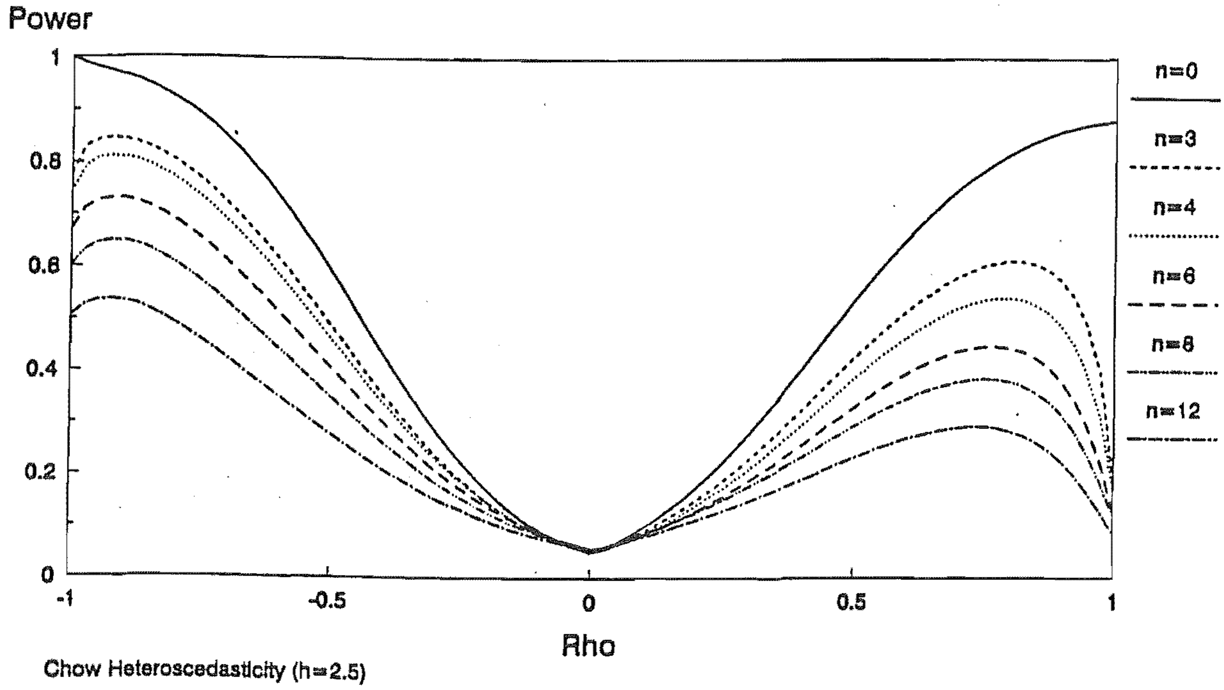


Figure 6.4b
Uniform Data; T = 20
S(0.5) Test

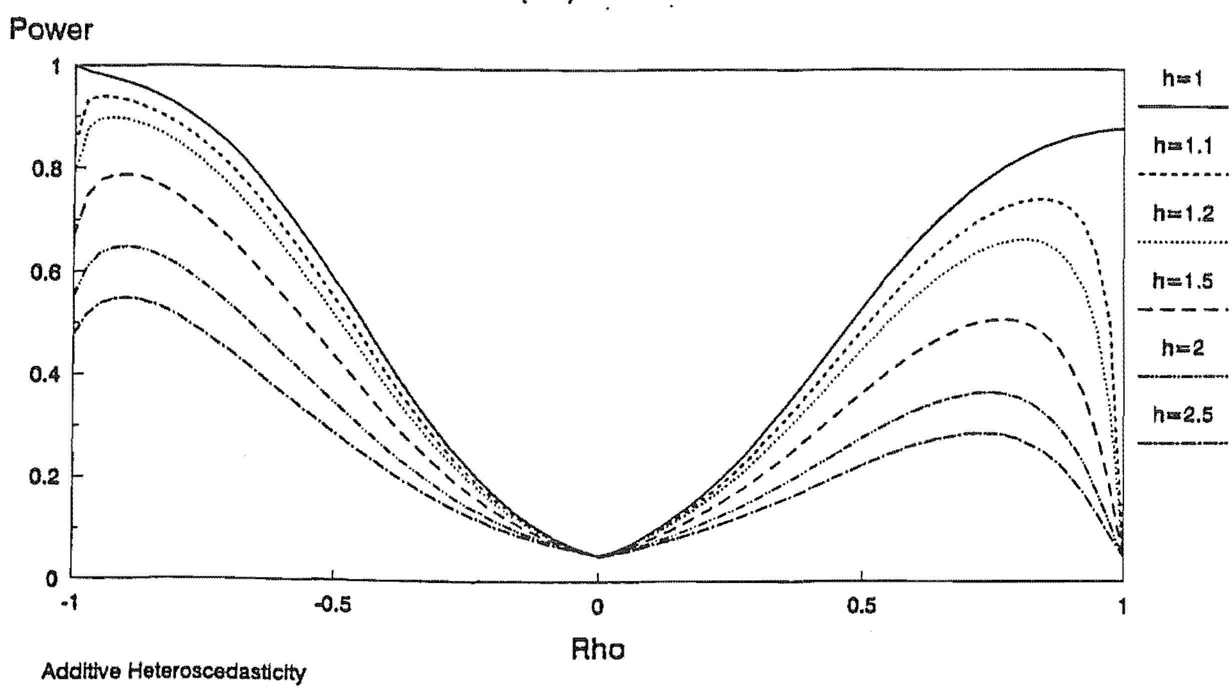


Figure 6.5a

Normal Data; T = 20

h = 1.0

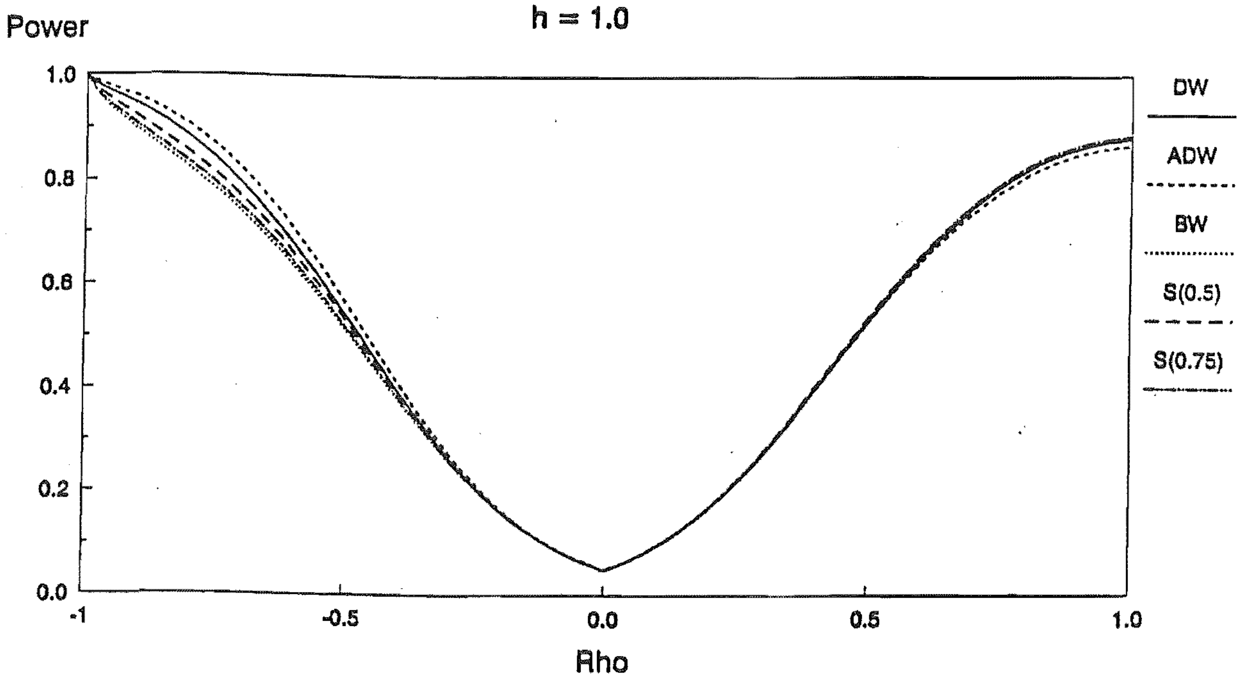
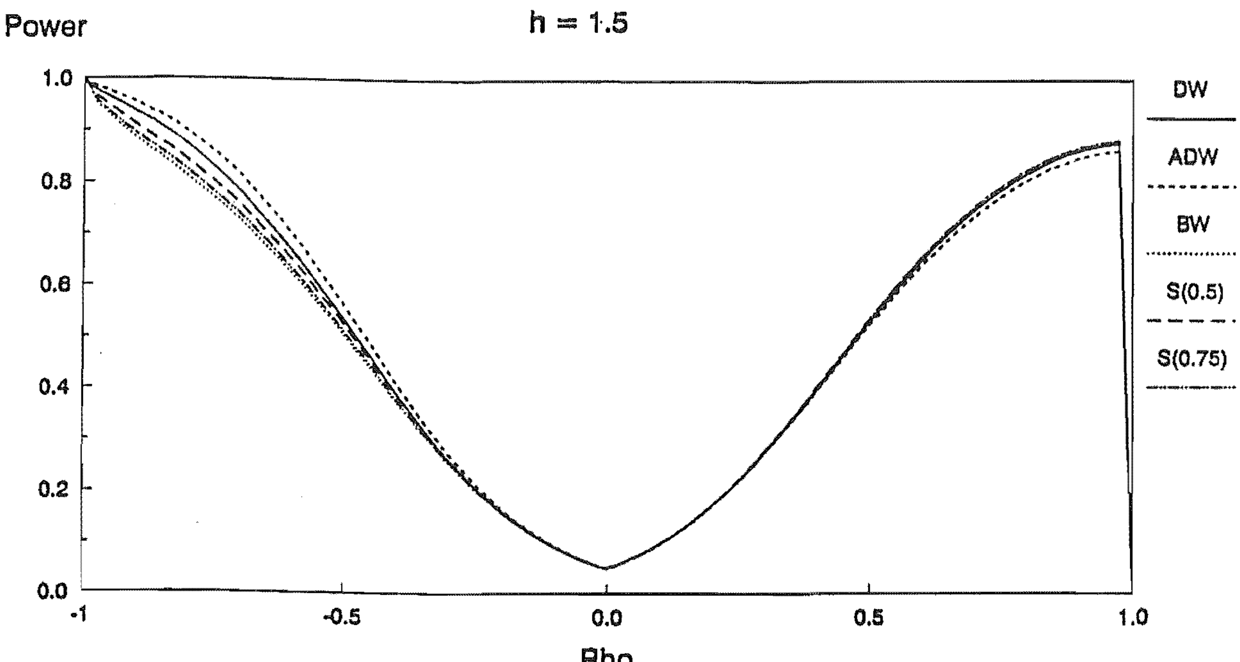


Figure 6.5b

Normal Data; T = 20

h = 1.5



Multiplicative Heteroscedasticity

Figure 6.6a

Spirits Data; T=20

DW Test

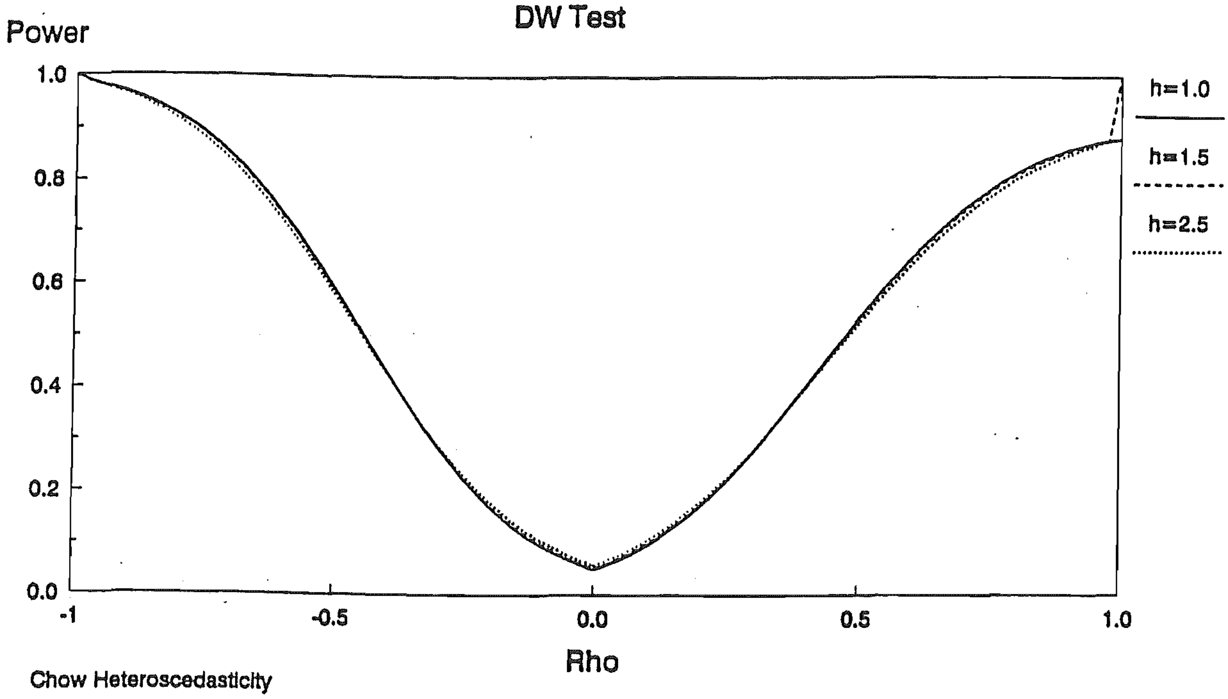
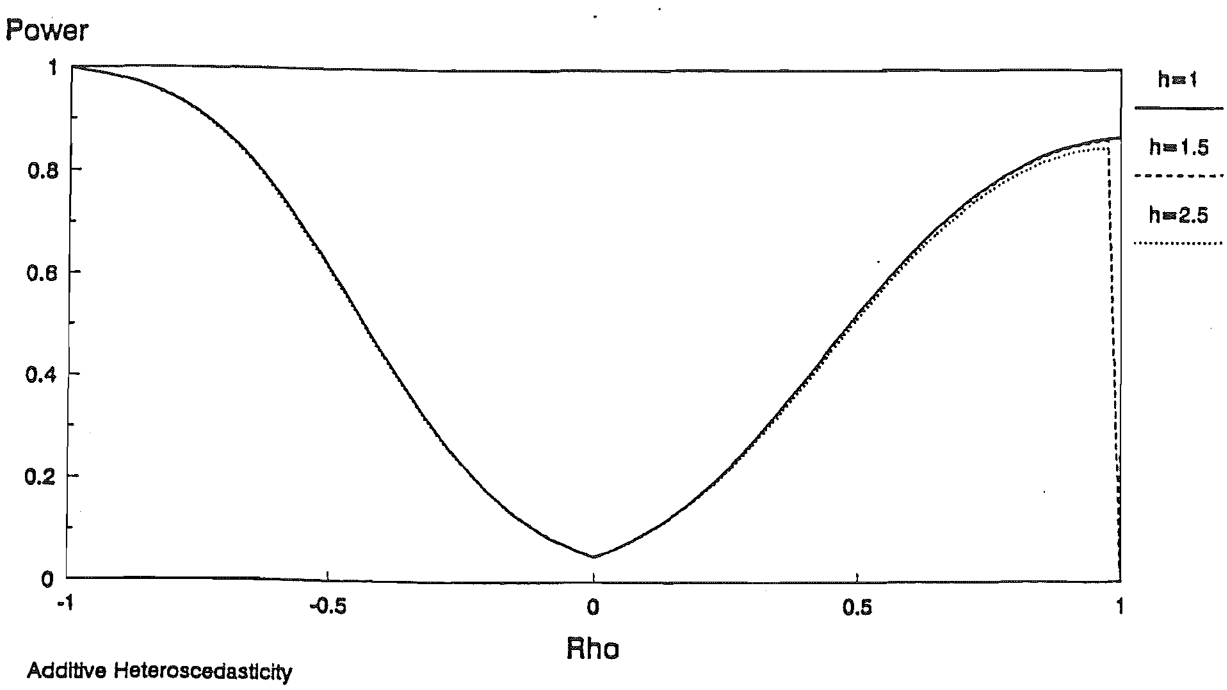


Figure 6.6b

Uniform Data; T=20

ADW Test



CHAPTER 7

TESTING FOR SERIAL INDEPENDENCE AGAINST INCORRECT ALTERNATIVES

7.1 Introduction

It was observed in chapters 3 and 4 that the rejection of the null hypothesis of serial independence on the basis of an autocorrelation test does not provide the researcher with reliable information about the precise nature of the true error process. The equivalence of LM tests against AR and MA processes of the same order, and the approximately LBI property of the DW test in the presence of MA(1) errors, are two examples which support this general contention.

In the present chapter we enlarge on this theme and provide empirical evidence of the sometimes very serious consequences of mis-specifying the alternative hypothesis for an autocorrelation test. All of the results reported in this chapter are exact in the sense of section 2.4; the cost of imposing this restriction is that tests which are only asymptotically justified are not considered. Throughout the chapter we confine our attention to exact AR(1) tests such as those used in chapter 6.

In particular, we consider the power of the DW, ADW BW and s(0.5) tests under a variety of data conditions. The design matrices used include the X1 to X6 matrices of chapter 6 and one additional matrix defined as

X7 A constant, a linear time trend and the logarithm of quarterly registered unemployed in New Zealand, commencing 1952(2).

This matrix was included for all error specifications considered in this chapter, although its major relevance is to the restricted AR(5) process used in section 7.4 below. The Unemployment data for the period chosen are strongly seasonal but without the very strong trend evident in more recent values of the series.

For each of the processes considered, a full investigation of the entire stationary parameter space for each test was conducted using a sample size of 20. A less complete series of evaluations used a sample size of 60 with the aim of highlighting the effect of more observations on the mis-specification consequences for these tests. It is known, in general, that larger samples produce larger powers (since all tests used are consistent tests) and that the power differences between the tests are reduced as the sample size grows. Thus, the major reason for concentrating on a relatively small sample size is that differences between tests are more easily discernible, as are the effects of mis-specification.

The remainder of the chapter is divided into three subsections (7.2 to 7.4) which consider the properties of the group of autocorrelation tests listed above when the true error process is, respectively, AR(2), MA(1) and a particular restricted AR(5) process defined by

$$u_t(1-\phi_1 L)(1-\phi_4 L^4) = \varepsilon_t \ , \text{ where } \varepsilon_t \sim N(0,\sigma_\varepsilon^2).$$

Some concluding comments are made in section 7.5.

In each case outlined above the assumed true process constitutes a mis-specification of the alternative hypothesis. The specifications chosen have a maximum of two parameters, giving the advantage of relative simplicity. They also have the ability to encompass other models as special cases; the restricted AR(5) model is a simple AR(4) process when $\phi_1=0$, for example.

Previous work which relates to the topics considered here is limited to the studies of King (1983), Weber and Monarchi (1982), Smith (1976) and Blattberg (1973) all of which evaluated the power of the DW test in the presence of a variety of error specifications. To the best of our knowledge, no authors have made a comparative study of the effects of MA(1) mis-specification on a group of AR(1) tests, and neither have the AR(2) and restricted AR(5) processes been considered in this context.

In each of the next three sections, we first discuss theoretical issues of relevance and then proceed to a summary of the results of the numerical computations undertaken. We omit discussion of the data, which has been covered in the previous chapter and earlier in this section. Similarly, the method of computing the power functions was detailed in chapter 6 and will not be repeated here.

7.2 Testing against AR(1) in the Presence of AR(2) Errors

In this section we consider the consequences of using tests, which are known to be powerful against AR(1) errors, in models for which the true error process is AR(2). Some theoretical characteristics of test power functions in such models are discussed in the next subsection. This is followed by a report of the findings of a set of numerical power evaluations which demonstrate that the power effects depend, among other factors, on the type of data used. It is also shown that test power does not necessarily increase with the level of the first autocorrelation coefficient.

7.2.1 Theoretical Issues

We shall consider the following model

(1) $$y_t = x_t'\beta + u_t$$

(2) $$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t$$

where y_t is the time t realisation of the dependent variable, x_t is a vector of time t observations on k regressors, β is a kx1 vector of parameters, ϕ_i, i=1,2 are autoregressive parameters and $\varepsilon_t \sim NID(0,\sigma_t^2)$. We suppose that a researcher applies a test of $H_0:\phi_1=0$ to the model, unaware of the possible presence of a non-zero ϕ_2.

Assuming that the variables in (1) have been appropriately differenced where necessary, attention is restricted to covariance stationary error processes. Goldberg (1958, p171) shows that this can be achieved by imposing the following constraints on the parameters ϕ_i, i=1,2.

(3) $$\phi_1 + \phi_2 < 1$$

(4) $$-\phi_1 + \phi_2 < 1$$

(5) $$\phi_2 > -1 \ .$$

Under these conditions the covariance matrix of u can be derived through the use of the Yule-Walker equations which, for this problem are:

(6) $$\rho(0) = \phi_1\rho(-1) + \phi_2\rho(-2)$$

(7) $$\rho(1) = \phi_1\rho(0) + \phi_2\rho(-1)$$

(8) $$\rho(\tau) = \phi_1\rho(\tau-1) + \phi_2\rho(\tau-2), \quad \text{for } \tau \geq 2.$$

Here $\rho(\tau)$ denotes the autocorrelation of u at lag τ. Recalling that $\rho(\tau) = \rho(-\tau)$ and that $\rho(0)=1$, this system can be solved to yield $\rho(1)=\phi_1/(1-\phi_2)$. Using the recursion of (8), the covariance matrix of an AR(2) process can now be constructed conditional on the sample size and the assumed values of the ϕ_i. Furthermore the scale factor σ_u^2, though irrelevant for the power of the tests considered here, can be derived by this method and is given by

$$\sigma_u^2 = \left(\frac{1-\phi_2}{1+\phi_2}\right)\frac{\sigma_\varepsilon^2}{[(1-\phi_2)^2-\phi_1^2]} \ ,$$

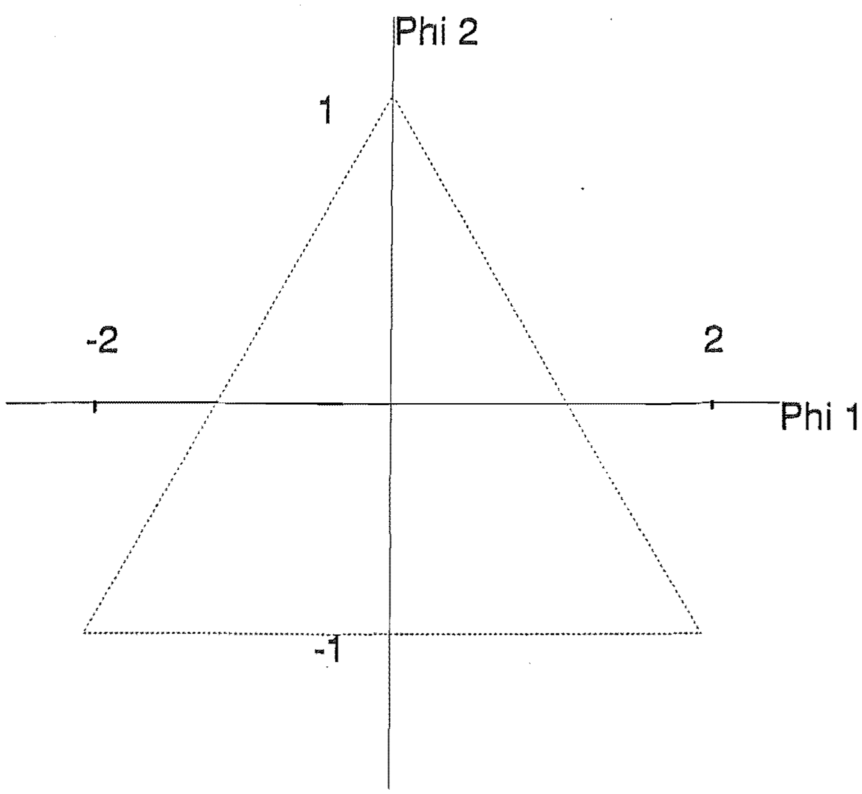
which can immediately be seen to collapse to the well known AR(1) scale factor when $\phi_2=0$. The covariance matrix of this process is a Toeplitz matrix with the first column being:

$$V_1 = \begin{bmatrix} 1 \\ \phi_1/(1-\phi_2) \\ (\phi_1^2+\phi_2-\phi_2^2)/(1-\phi_2) \\ (\phi_1^3+2\phi_1\phi_2-\phi_1\phi_2^2)/(1-\phi_2) \\ (\phi_1^4+3\phi_1^2\phi_2-\phi_1^2\phi_2^2+\phi_2^2-\phi_2^3)/(1-\phi_2) \\ (\phi_1^5+4\phi_1^3\phi_2-\phi_1^3\phi_2^2+3\phi_1\phi_2^2-2\phi_1\phi_2^3)/(1-\phi_2) \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} .$$

The stationarity conditions define a two dimensional region in ϕ_1, ϕ_2 space which is most easily visualised as being contained within the dashed lines in Figure 7.2.1.

Figure 7.2.1

Stationarity Region for AR(2) Process



The dependence of two boundaries of the stationarity region on both of the ϕ_i makes the analysis of limiting powers for these lines particularly troublesome. The boundary defined by $\phi_2 = -1$ can, however, be handled using the methods developed in the previous chapter. When $\phi_2 = -1$ the first column of the covariance matrix of u is given by:

$$V(\phi_1,-1)_1 = \begin{bmatrix} 1 \\ \phi_1/2 \\ \phi_1^2/2 - 1 \\ \phi_1^3/2 - 3\phi_1/2 \\ \phi_1^4/2 - 2\phi_1^2 + 1 \\ \phi_1^5/2 - 5\phi_1^3/2 + 5\phi_1/2 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}.$$

In general $V(\phi_1,-1)$ is a full rank matrix so that the exact powers of the various tests can be computed with the standard algorithms using T-k non-zero eigenvalues. At the endpoints of the $\phi_2 = -1$ boundary line the rank of V is severely reduced, however. Using the expression for $V(\phi_1,-1)_1$ above it can be shown that when $\phi_1=1$ the autocorrelation function of u follows the pattern:

$$\rho_0 = 1$$
$$\rho_1 = 1/2$$
$$\rho_2 = -1/2$$
$$\rho_3 = -1$$
$$\rho_4 = -1/2$$
$$\rho_5 = 1/2$$
$$\rho_6 = 1$$

which repeats indefinitely. Thus the rank of V(1,-1) is six and the power of each test depends on the signs of the six non-zero eigenvalues of (Q - r*M)V. The situation is somewhat more severe at the other end of the ϕ_2=-1 boundary, where ϕ_1=-1. In this case the individual autocorrelations can only take the values 1 or -1/2 and do so in the following pattern:

$$\rho_0 = 1$$
$$\rho_1 = -1/2$$
$$\rho_2 = -1/2$$
$$\rho_3 = 1$$
$$\rho_4 = -1/2$$
$$\rho_5 = -1/2$$
$$\rho_6 = 1, \textit{ etc}$$

It is clear that at this corner of the stationarity region boundary the power of each test depends upon only three non-zero eigenvalues, since that is the rank of V(-1,-1). The likelihood of obtaining a deterministic power value of either zero or unity (corresponding to cases in which each of the non-zero eigenvalues has a positive or negative sign respectively) is clearly greater, the smaller is the rank of V.

7.2.2 Numerical Evaluations

For this study power functions were calculated for a grid of ϕ_1 values with ϕ_2 taking the fixed values of 0, 0.2, 0.4, 0.6, and 0.8. The direction was then reversed and power functions for

$\phi_1=(0,\ 0.2,\ 0.4,\ 0.6,\ 0.8)$ were evaluated for a grid of ϕ_2's. In each direction the power calculations were restricted to the stationary region and points on the boundary of this region. Tests of $H_0:\phi_1=0$ were conducted against the one sided positive alternative $H_a^+:\phi_1>0$. Each of the matrices X1 to X7 inclusive was employed in a thorough investigation of test power with a sample of 20 observations.

There are several interesting findings which result from the computation of power functions for these tests. We begin with a discussion of the sizes of the tests, which were found to be data dependent for a given parameter set. All Tables and Figures referred to are located at the end of section 7.2.

For the X1 to X5 design matrices the sizes of the tests follow a similar pattern. From a nominal (correctly specified) size of 0.05 the true sizes of all tests with these data initially increase slightly with ϕ_2 to values in the region of 0.06 when $\phi_2=0.3$. Sizes then decline to an average of 0.03 when $\phi_2=0.8$ and finally to zero, in the limit as $\phi_2\rightarrow 1$. For negative ϕ_2 the true size, using these data, always declines with $|\phi_2|$. The solid line in Figure 7.2.2 shows how the size of the ADW test changes with ϕ_2 when the X1 matrix is used, while Tables 7.2.1 and 7.2.2 give rejection frequencies for the same test in the context of the CPI data (X2).

For the X7 matrix a similar pattern was evident but without the initial increase in test size (see Tables 7.2.3 and 7.2.4 for BW powers using these data). Consequently the size of all tests along the ϕ_2=0.8 line was somewhat lower, averaging only 0.019. The size distortions found with the strongly seasonal X7 matrix are reasonably large, in both relative and absolute terms. They are minor, however, compared with the size distortion encountered with the X6 matrix when the BW and s(ρ_1) tests were used. In these cases the true size of the test was always greater than the nominal size for $\phi_2 \neq 0$, ranging from 0.09 (for ϕ_2=0.2 with the s(0.5) test) to 0.47 (for the BW test as $\phi_2 \rightarrow 1$). The solid line in Figure 7.2.3 illustrates the severity of the size distortion using the BW test and the X6 matrix.

One of the major justifications for the use of the BW and s(ρ_1) tests has been their high power in the context of the X6 design matrix. This has been established empirically by King (1985) but, apart from Evans (1992) and the new work reported in this thesis, the robustness of these tests to various forms of mis-specification has not been examined. The size distortion discussed above reveals a major weakness of these tests compared to the more standard DW and ADW tests. Clearly high power is of little value in tests exhibiting such sharp increases above nominal size.

With the above size distortions in mind, we turn now to a discussion of the powers of the tests when the true errors are AR(2). For this discussion a distinction is made between the two components of the AR(2) process. From the autocorrelation functions presented in section 7.2.1 it is clear that both ϕ_1 and ϕ_2 determine the size of the first order autocorrelation coefficient. We will treat these components separately, however, using ϕ_1 as the parameter which indexes the power function and ϕ_2 as the degree of mis-specification.

The effects of ϕ_2 on the powers of the various tests can best be discussed in two groups. The first group includes all design matrices and tests with the exception of the X6 matrix with the BW and $s(\rho_1)$ tests. These tests used in the context of the X6 matrix constitute the second group, for which different power effects were found. Notice that the division between these groups directly parallels the size distortion results, as would be expected.

Figure 7.2.4 shows the power of the DW test with the X7 matrix and is typical of the first group. It should be noted that these power functions are computed only for stationary parameter values, hence the larger is ϕ_2, the fewer values of ϕ_1 are permissible. This graph shows that for moderate values (*i.e.*, no greater than 0.4) of both ϕ_1 and ϕ_2 there is relatively little effect on the power of the AR(1) test. For larger values of either parameter, however, the power effects of mis-specification are more notable. In particular, the power loss, relative to the correctly specified model, is more severe as the non-stationarity boundary is approached. Extreme examples of power loss occur with the DW and ADW tests in conjunction with Watson's matrix (X6). One such case is shown in Figure 7.2.6 where the power of the ADW test, which is not monotonic in ϕ_1 even in a correctly specified model (depicted by the solid line), is shown to be further reduced by the presence of a non-zero ϕ_2.

The second group is represented by Figure 7.2.5 which shows the power of the s(0.5) test with Watson's (X6) matrix. The upward size distortion typical of this group is clearly evident in this Figure. The relative "advantage" of greater size is not maintained as ϕ_1 increases, however, as is revealed by the tendency for the power curves to converge towards the correctly specified (ϕ_2=0) line. This pattern is similar to that noted for group one, in that as the non-stationarity boundary is approached the relative power gain is reduced. As noted above, the major concern with the "group two" cases is the greatly increased sizes induced by mis-specification of this type. Figure 7.2.5 also reveals that rejection of the null in this model may be caused by negative values of ϕ_1, even though the test is against H_a:ρ>0. Again this is typical of the group two tests.

Blattberg (1973), in an early study of the DW test against AR(2) errors, predicted on the basis of asymptotic analysis that the power of the test would increase (decrease) as the first autocorrelation coefficient, $\rho_1 = \dfrac{\phi_1}{1-\phi_2}$, became more positive (negative). This is not a general result, however, as can be seen by looking down the columns of Table 7.2.4, for example, where exactly the reverse phenomenon occurs. In this study the only cases which support such a relation between power and ρ_1 are those using Watson's (X6) matrix with the BW (see Figure 7.2.3) and $s(\rho_1)$ tests[1].

7.2.3 Conclusion

To conclude this section we reiterate the major findings. First, the effect of AR(2) errors on the size and power of the DW and ADW tests is relatively minor for parameter values which are not near the non-stationarity boundary. When the parameters do not satisfy this condition, however, reasonably severe power losses occur. Second, these results also apply to the Kadiyala based BW and $s(\rho_1)$ tests for all data except X6. Third, the true sizes of the Kadiyala based tests with Watson's (X6) matrix are dramatically increased above their nominal levels by the addition of the ϕ_2 component. The power functions in these cases, however, converge towards the correctly specified power as the parameters approach non-stationarity so that the effect of the mis-specification is reduced. Finally, for the latter group of tests, negative ϕ_1's combined with positive ϕ_2's can lead to rejection of the null against a positive one-sided alternative.

---

[1] Blattberg's own empirical work did not generally support the proposition either.

TABLE 7.2.1
Power of ADW Test with AR(2) Errors
CPI Data (X2)

	$\phi 2$				
$\phi 1$	0.0	0.2	0.4	0.6	0.8
-0.9	0.000				
-0.7	0.000	0.000			
-0.5	0.001	0.001	0.000		
-0.3	0.004	0.005	0.004	0.003	
-0.1	0.023	0.026	0.025	0.021	0.011
0.0	0.050	0.054	0.052	0.044	0.028
0.1	0.098	0.102	0.098	0.083	0.055
0.3	0.276	0.271	0.250	0.204	
0.5	0.523	0.490	0.429		
0.7	0.726	0.661			
0.9	0.837				

TABLE 7.2.2
Power of ADW Test with AR(2) Errors
CPI Data (X2)

	$\phi 1$				
$\phi 2$	0.0	0.2	0.4	0.6	0.8
-1.0	0.000	0.000	0.000	1.000	1.000
-0.9	0.002	0.025	0.255	0.776	0.959
-0.7	0.010	0.076	0.345	0.723	0.916
-0.5	0.023	0.120	0.382	0.698	0.885
-0.3	0.036	0.151	0.399	0.677	0.855
-0.1	0.046	0.169	0.401	0.652	0.819
0.0	0.050	0.173	0.398	0.635	0.793
0.1	0.053	0.176	0.391	0.614	0.758
0.3	0.054	0.171	0.365	0.549	
0.5	0.049	0.154	0.311		
0.7	0.037	0.117			
0.9	0.017				
1.0	0.000				

TABLE 7.2.3 Power of BW Test with AR(2) Errors Unemployment Data (X7)					
	$\phi 2$				
$\phi 1$	0.0	0.2	0.4	0.6	0.8
-0.9	0.000				
-0.7	0.000	0.000			
-0.5	0.001	0.001	0.000		
-0.3	0.005	0.005	0.004	0.002	
-0.1	0.025	0.023	0.021	0.016	0.008
0.0	0.050	0.048	0.043	0.034	0.020
0.1	0.094	0.090	0.081	0.065	0.040
0.3	0.258	0.244	0.217	0.166	
0.5	0.497	0.460	0.390		
0.7	0.712	0.641			
0.9	0.836				

TABLE 7.2.4
Power of BW Test with AR(2) Errors
Unemployment Data (X7)

	$\phi 1$				
$\phi 2$	0.0	0.2	0.4	0.6	0.8
-1.0	0.388	1.000	1.000	1.000	1.000
-0.9	0.179	0.240	0.362	0.758	0.954
-0.7	0.085	0.185	0.396	0.712	0.908
-0.5	0.060	0.169	0.397	0.684	0.875
-0.3	0.053	0.166	0.391	0.658	0.847
-0.1	0.051	0.164	0.381	0.631	0.812
0.0	0.050	0.163	0.373	0.614	0.787
0.1	0.049	0.160	0.363	0.592	0.749
0.3	0.046	0.149	0.331	0.518	
0.5	0.039	0.127	0.269		
0.7	0.027	0.090			
0.9	0.012				
1.0	0.000				

Figure 7.2.2
Power of ADW Test against AR(2) Errors
Spirits Data; T=20; 5% Size

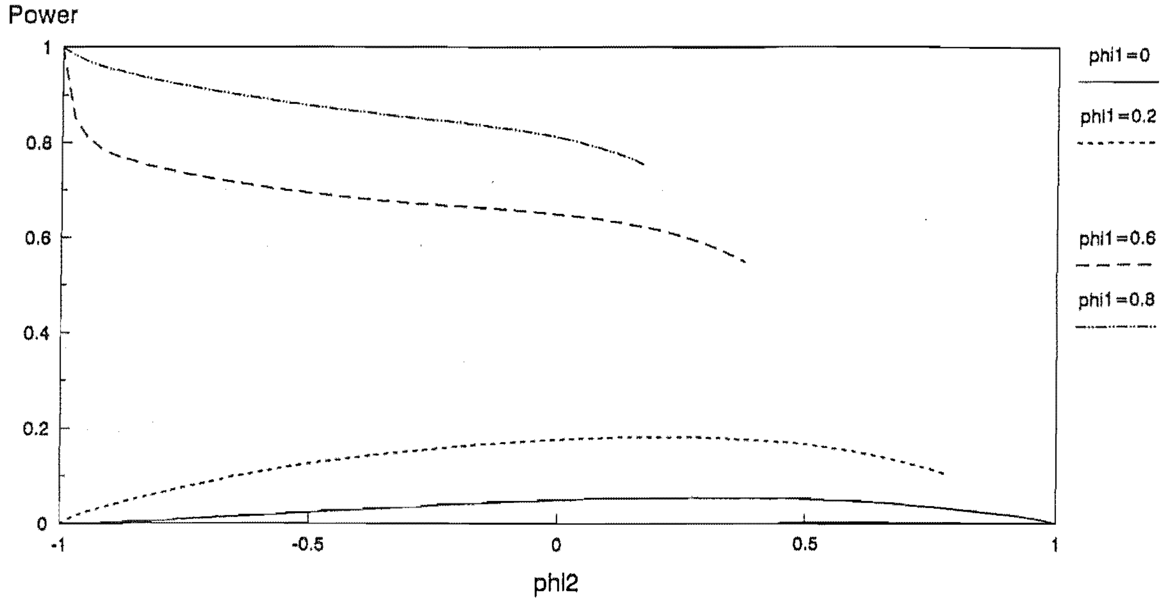


Figure 7.2.3
Power of BW Test against AR(2) Errors
Watson's Data; T=20; 5% Size

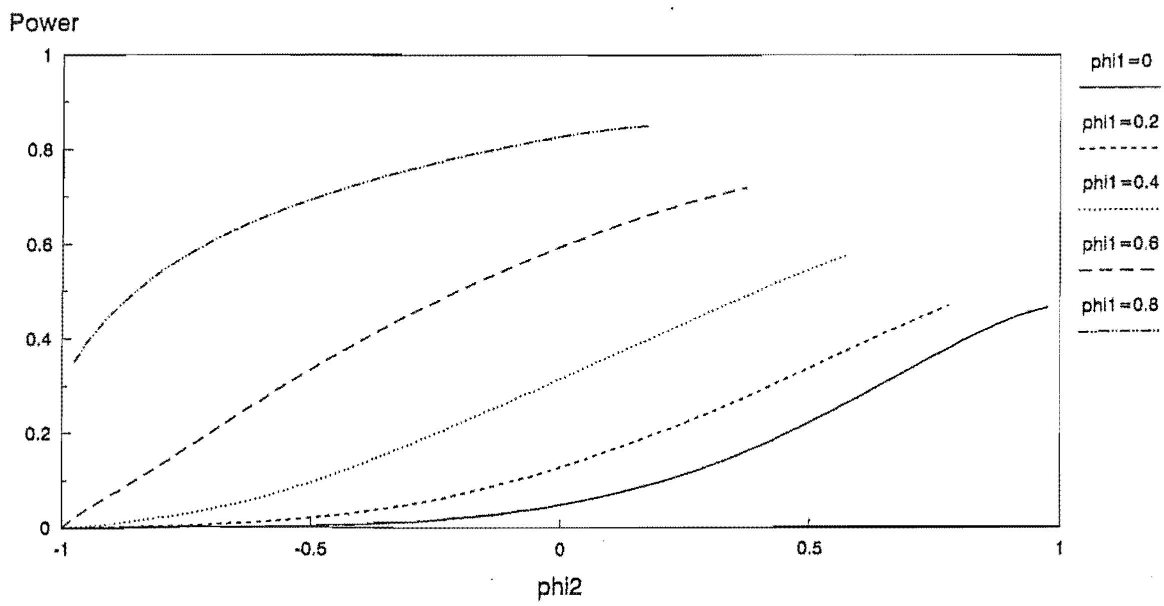


Figure 7.2.4

Power of DW Test against AR(2) Errors
Unemployment Data; T=20; 5% size

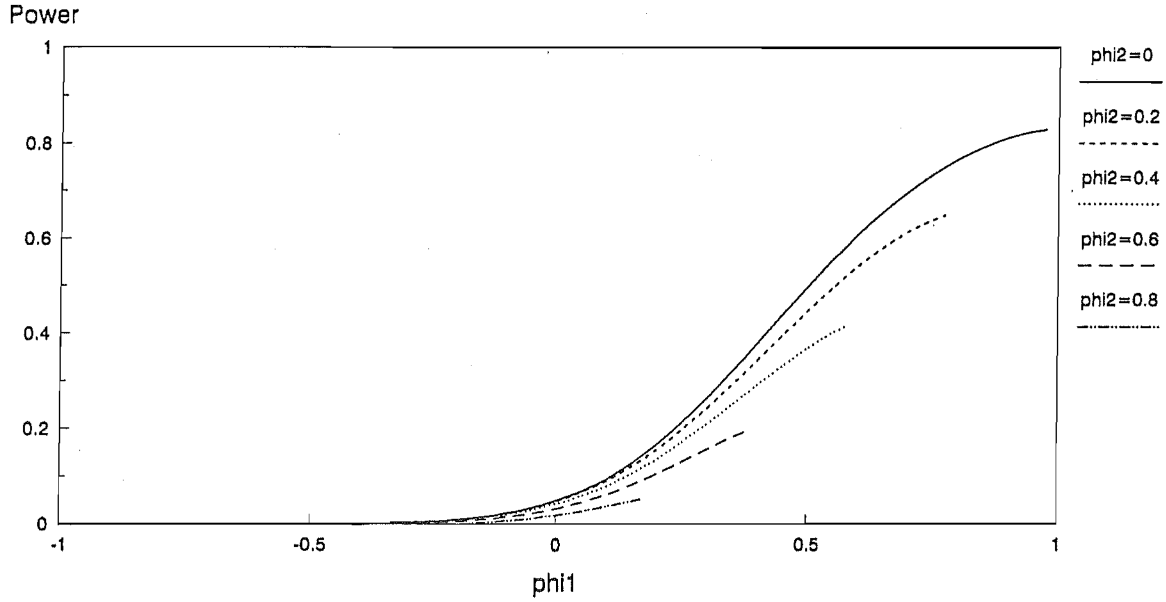


Figure 7.2.5

Power of s(0.5) Test against AR(2) Errors
Watson's Data; T=20; 5% Size

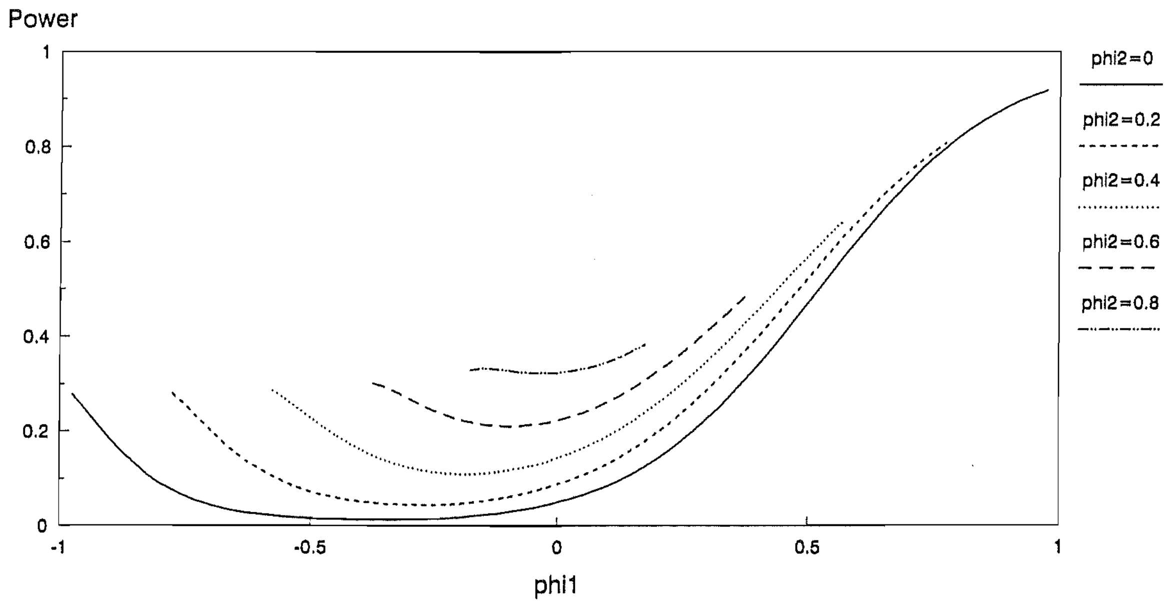


Figure 7.2.6

Power of ADW Test against AR(2) Errors

Watson's Data; T=20; 5% size

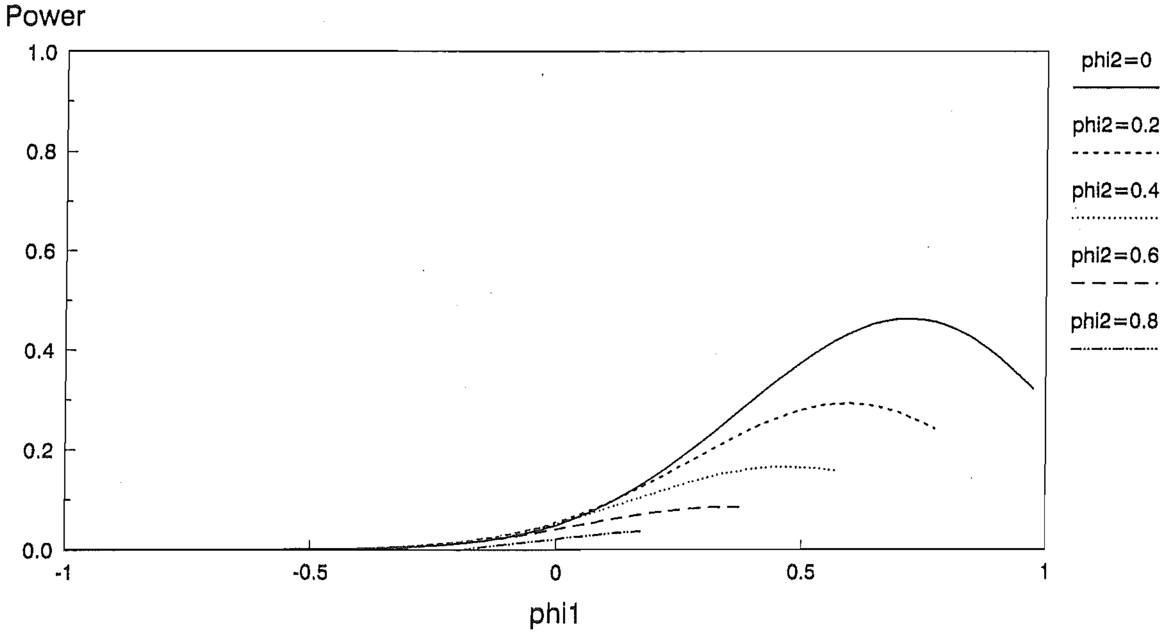
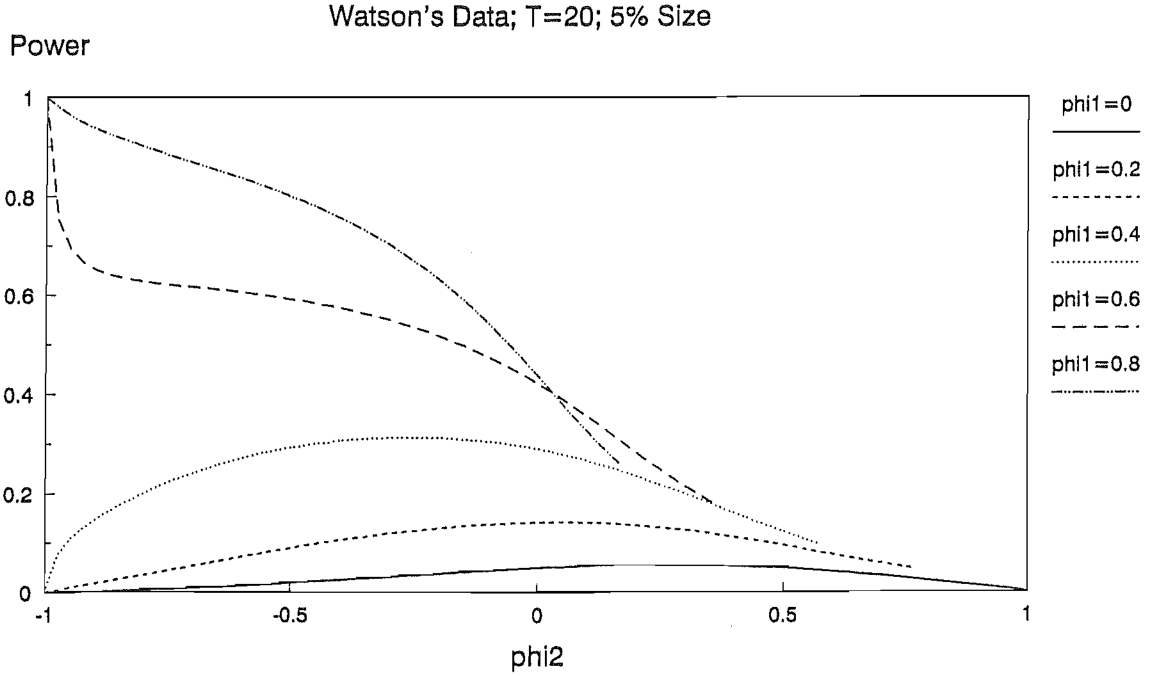


Figure 7.2.7

Power of DW Test against AR(2) Errors

Watson's Data; T=20; 5% Size



7.3 Testing against AR(1) in the Presence of MA(1) Errors

Moving average error specifications are generated naturally in a variety of economic contexts. A good example is the adaptive expectations framework. The estimation of moving average models is not trivial (as is the case for AR processes) unless quite strong assumptions are made about the initial value of the process[2]. This fact may account for the relatively light emphasis on such models in the hypothesis testing literature.

In this section we evaluate the ability of a group of AR(1) tests to reject serial independence when the true errors are MA(1). The implications of high power here are rather different from those of section 7.2 as the AR(1) process is not nested by the model being employed. Thus, high rejection rates, which signal significant autocorrelation, may well be interpreted as being indicative of AR(1) errors, rather than the true MA(1) process. A standard correction for AR(1) errors, such as a simple Cochrane-Orcutt transformation for example, will not produce a scalar covariance matrix for the regression errors in this case. Autocorrelated errors will remain with the attendant problems of inefficient parameter estimates and incorrect standard errors.

The next two subsections discuss the model and the empirical study. These are followed by some concluding comments.

[2]Assuming the first value is fixed at zero allows estimation via a simple Gauss-Newton algorithm.

7.3.1 Theoretical Issues

The model used in this section is described by (1) with

$$u_t = \varepsilon_t + \theta\varepsilon_{t-1}\ ,\ \varepsilon_t \sim N(0,\sigma_\varepsilon^2)\ . \tag{9}$$

This first order moving average process has a variance of $\sigma_u^2 = \sigma_\varepsilon^2(1+\theta^2)$ and an autocorrelation function in which the first autocorrelation, $\rho(1) = \theta/(1+\theta^2)$ while $\rho(\tau)=0$ for all $\tau \geq 2$. It is easily shown that an MA(q) process is always stationary. It is, however, customary to impose the invertibility constraint, $|\theta|<1$, on the MA(1) process. This simply ensures that only one parameter is consistent with a given autocorrelation function. In the absence of this constraint, for example, it is clear that $\varphi = 1/\theta$ produces an exactly identical autocorrelation function to that generated by the parameter θ. It should be noted that imposing the invertibility constraint limits the size of the first autocorrelation to be less than 1/2 in absolute value.

To elaborate briefly on the consequences of erroneously concluding that the errors from the regression in (1) are given by (2) rather than (9), let us consider the effect of applying a Cochrane-Orcutt transformation to (9). Assuming the AR(1) parameter to be ϕ, the regression error term becomes

$$\begin{aligned} u_t^* &= u_t - \phi u_{t-1} \\ &= (\varepsilon_t + \theta\varepsilon_{t-1}) - \phi(\varepsilon_{t-1} + \theta\varepsilon_{t-2}) \\ &= \varepsilon_t + (\theta-\phi)\ \varepsilon_{t-1} - \theta\varepsilon_{t-2}\ . \end{aligned}$$

The inappropriate use of this transformation in the presence of MA(1) errors will therefore produce an MA(2) error process with no first order component in the special case that the assumed ϕ is identical to the true θ.

Given the characteristics of the autocorrelation function described above some initial conjectures about the effects of MA(1) errors can be made. The most important distinguishing feature of the MA(1) process is the absence of a long term memory. This, of course, is its major attraction for cases (such as the adaptive expectations model) in which the error in a forecast affects the next period only. Simply because of this truncation of the autocorrelation function we would expect a test designed to detect AR processes to have less power against MA alternatives of the same order.

Even abstracting from the effect of truncation, however, we can reasonably expect that the difference in power (against AR(1) and MA(1) processes with the same parameter size) will be greater, the larger are the parameters involved. The reason for this is that when $\phi = \theta \neq 0$ (where ϕ is the autoregressive parameter in an AR(1) model) the $\rho(1)$ from the MA model is smaller than that from the AR model, and the difference increases with ϕ,θ. This finding should be kept in mind throughout this section when comparisons are made between AR(1) and MA(1) model with the same parameter values.

King (1983) shows that the DW test is an approximately LBI test of $H_0{:}\theta=0$ while the ADW test is truly LBI for this problem. Introducing a point optimal test which is LBI at a selected value of θ, King shows that this latter test can have significantly greater power against moderate to large values of θ, relative to the power acheived by either the DW or ADW tests. In this section we continue to restrict attention to AR(1) tests, but widen the range of such tests beyond those considered by King (1983) as well as using several different design matrices. X1 and X2 were both used by King but none of X3 to X7 were experimented with.

7.3.2 Numerical Evaluations

For this study the group of tests for serial independence, which was described at the beginning of the Chapter, were applied for both $H_a^+:\phi>0$ and $H_a^-:\phi<0$. All tests were applied at a nominal 5% significance level and the limiting powers as the non-stationarity boundary is approached (in either direction) were computed by the method of Krämer and Zeisel (1990) which was described in chapter 6. It is appropriate to begin the discussion of the effects of the MA(1) mis-specification by considering the consequences for the sizes of the tests. The Tables and Figures referred to in this subsection may be found at the end of section 7.3.

When the true θ in (9) is zero, the covariance matrix of u is a scalar matrix, just as it would be under the null of an AR(1) model. For this reason, we would expect the degree of size distortion induced by MA(1) errors to be literally zero. This was found in the numerical evaluations as can be seen in Tables 7.3.1 to 7.3.4 inclusive. Thus direct comparisons can be made validly between the AR(1) and MA(1) power functions.

The power comparisons using all data matrices except X6 produced very similar conclusions. When the true errors are MA(1), rather than AR(1), the powers of the AR(1) tests are considerably reduced for large (greater than 0.6) positive values of θ. Table 7.3.1 shows the power of the DW and BW tests for both AR(1) and MA(1) errors when the data matrix is X3. These data provide some of the more pronounced power differences, with both tests showing 23% less power against θ=0.7 than when ϕ=0.7. Against H_a^-, serious power reduction, relative to AR(1) errors, began at smaller absolute values of θ (such as θ = -0.3). Figures 7.3.1 and 7.3.2 support these statements and are typical of the power functions of all tests evaluated for these data. For a comparison across tests, Figure 7.3.3 shows that, of the tests used here, the ADW test dominates all of the others considered against both positive and negative alternatives. This figure is representative of all of the data matrices except X6.[3]

For the X6 matrix it is well known that the power functions of the DW and ADW tests are not monotonic in the degree of autocorrelation, even in correctly specified models. A major justification for the Kadiyala-based point optimal tests is that they have power functions which retain the desirable monotonicity property even when used with extreme data such as X6. Figure

[3] It should be noted, however, that the point optimal test introduced by King (1983) is superior to any considered here for this model, and should be used in cases where the possibility of MA(1) errors is explicitly entertained.

7.3.4 shows the effect of both AR(1) and MA(1) errors on the power of the DW test. The dramatic downturn of the power function against AR(1) errors is clearly evident. Interestingly, this feature is absent from the power against MA(1) errors, resulting in a higher DW power for large (absolute) values of θ, relative to power against similar values of ρ.

The case of the point optimal tests, used with Watson's X6 matrix, is illustrated in Figure 7.3.5, while Table 7.3.4 gives power values for the DW and BW tests[4] with this design matrix. Here we see that the power function against MA(1) errors is considerably lower, but has the same general shape, as was found for all tests with the first group of data sets. It is also apparent that the s(0.5) test is very powerful against a correctly specified AR(1) process with a positive parameter. The non-monotonic power of the s(0.5) test against H_a^-, which is evident in Figure 7.3.5, is not a cause for concern. This is because the s(0.5) test is designed only for tests against H_a^+ so it is somewhat unfair to use it in this manner[5]. Notwithstanding this, it should be noted that for all other data sets the point optimal tests performed better against a negative alternative hypothesis than against the positive alternatives for which they were designed. Seen in this light, it is surprising that the X6

[4] Recall from Chapter 3 that the BW test is point-optimal when ϕ=1.

[5] An alternative test, which is LBI when ρ = -0.5, can be constructed for tests against H_a^-.

matrix should have such a dramatic effect on the BW and $s(\rho_1)$ tests. To emphasise this point Figure 7.3.6 shows power functions for both the BW (representing the Kadiyala based tests) and the DW tests using X6 with both AR and MA errors. It is clear from this figure that the DW test is more powerful than the BW test against negative AR(1) errors, and also has a power advantage against both positive and negative MA(1) processes.

An interesting by-product of this study is the observation that in almost all cases (X6 is the only exception) the limiting power of an AR(1) test in a **correctly** specified model is unity as $\rho \to -1$, but this is never true (for the data used here) as $\rho \to 1$. This phenomenon can be clearly seen in Figures 7.3.1 and 7.3.2.

7.3.3 Conclusion

This section has shown that the tests against AR(1) errors investigated here are considerably less powerful against MA(1) errors than against AR(1) errors with the same parameter size. Our earlier conjecture, based on the structure of the autocorrelation functions, that the drop in power would be more pronounced the larger were the parameters, has been shown to be correct for all of the data sets that we have considered except X6. We have also reviewed some known results in connection with the X6 matrix and exposed a weakness in the BW and $s(\rho_1)$ tests, as compared to the DW and ADW tests, when testing against negative alternatives with this design matrix.

TABLE 7.3.1
Power of DW and BW Tests with MA(1) and AR(1) Errors
Normal Data (X3); 5% Size

	DW		BW	
ϕ, θ	MA(1)	AR(1)	MA(1)	AR(1)
-1.0	0.501	1.000	0.501	1.000
-0.9	0.497	0.942	0.497	0.893
-0.7	0.456	0.792	0.456	0.739
-0.5	0.359	0.545	0.359	0.516
-0.3	0.218	0.272	0.218	0.264
-0.1	0.092	0.095	0.092	0.094
0.0	0.050	0.050	0.050	0.050
0.1	0.094	0.096	0.094	0.096
0.3	0.247	0.273	0.247	0.273
0.5	0.436	0.530	0.436	0.534
0.7	0.575	0.749	0.575	0.755
0.9	0.633	0.863	0.633	0.869
1.0	0.639	0.881	0.639	0.886

TABLE 7.3.2
Power of DW and BW Tests with MA(1) and AR(1) Errors
Uniform Data (X4); 5% Size

	DW		BW	
ϕ, θ	MA(1)	AR(1)	MA(1)	AR(1)
-1.0	0.514	1.000	0.506	1.000
-0.9	0.510	0.962	0.502	0.955
-0.7	0.468	0.839	0.461	0.823
-0.5	0.369	0.585	0.364	0.571
-0.3	0.225	0.287	0.223	0.282
-0.1	0.093	0.096	0.093	0.096
0.0	0.050	0.050	0.050	0.050
0.1	0.096	0.096	0.095	0.096
0.3	0.258	0.273	0.257	0.273
0.5	0.467	0.528	0.466	0.534
0.7	0.621	0.747	0.620	0.755
0.9	0.685	0.864	0.685	0.870
1.0	0.691	0.881	0.691	0.887

TABLE 7.3.3
Power of DW and BW Tests with MA(1) and AR(1) Errors
Lognormal Data (X5); 5% Size

	DW		BW	
ϕ, θ	MA(1)	AR(1)	MA(1)	AR(1)
-1.0	0.490	1.000	0.484	1.000
-0.9	0.487	0.949	0.480	0.936
-0.7	0.448	0.817	0.442	0.798
-0.5	0.356	0.570	0.352	0.555
-0.3	0.220	0.282	0.218	0.277
-0.1	0.093	0.096	0.092	0.096
0.0	0.050	0.050	0.050	0.050
0.1	0.094	0.095	0.093	0.093
0.3	0.251	0.262	0.246	0.263
0.5	0.457	0.507	0.449	0.518
0.7	0.611	0.725	0.602	0.743
0.9	0.676	0.846	0.667	0.862
1.0	0.683	0.865	0.674	0.880

TABLE 7.3.4 Power of DW and BW Tests with MA(1) and AR(1) Errors Watson's Data (X6); 5% Size				
	DW		BW	
ϕ, θ	MA(1)	AR(1)	MA(1)	AR(1)
-1.0	0.412	0.000	0.364	1.000
-0.9	0.408	0.384	0.362	0.181
-0.7	0.376	0.461	0.335	0.294
-0.5	0.299	0.372	0.273	0.284
-0.3	0.189	0.214	0.178	0.188
-0.1	0.086	0.088	0.084	0.085
0.0	0.050	0.050	0.050	0.050
0.1	0.087	0.088	0.077	0.079
0.3	0.200	0.212	0.163	0.207
0.5	0.333	0.366	0.276	0.451
0.7	0.429	0.452	0.367	0.726
0.9	0.472	0.386	0.409	0.897
1.0	0.476	0.308	0.414	0.936

Figure 7.3.1
Power of ADW Test against MA(1) Errors
Unemployment Data; T=20; 5% Size

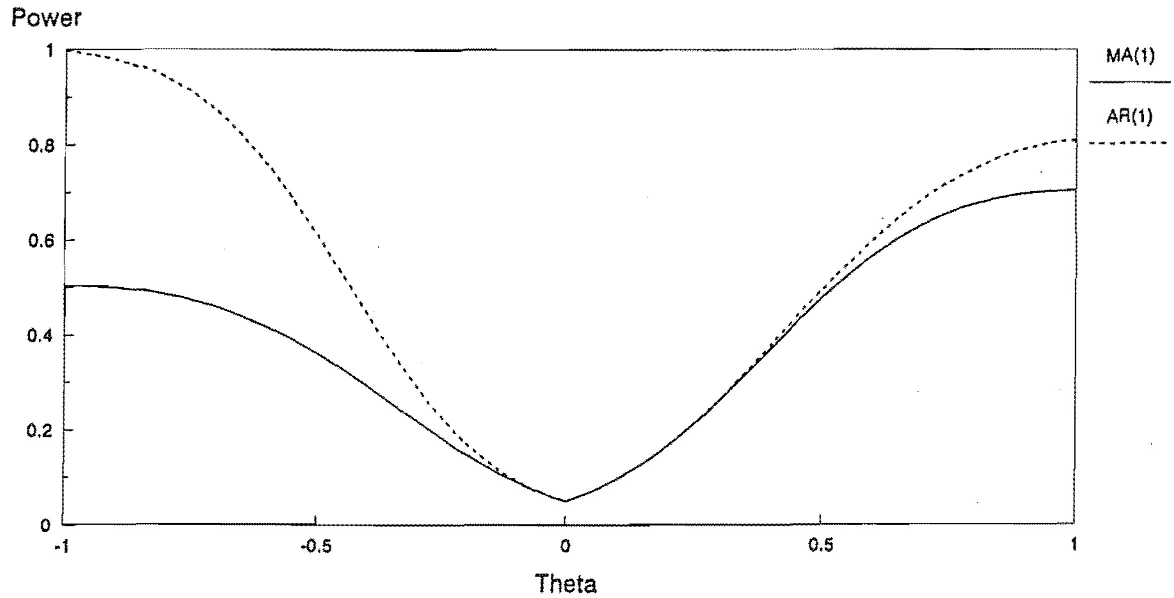


Figure 7.3.2
Power of BW Test against MA(1) errors
Uniform Data; T=20; 5% Size

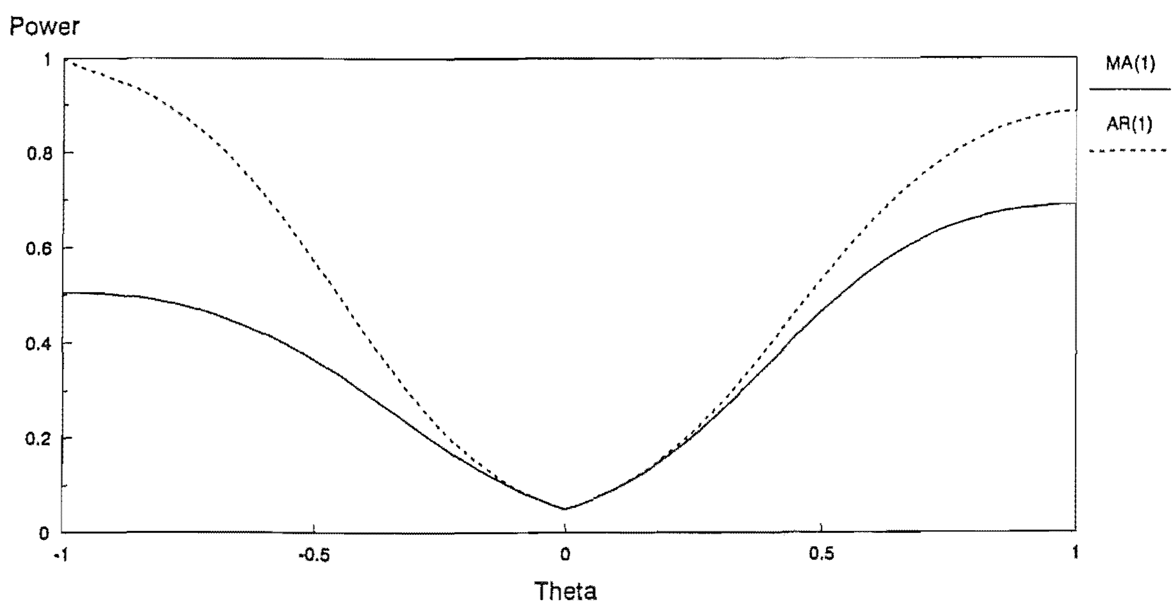


Figure 7.3.3
Power of All Tests against MA(1) Errors
CPI Data; T=20; 5% Size

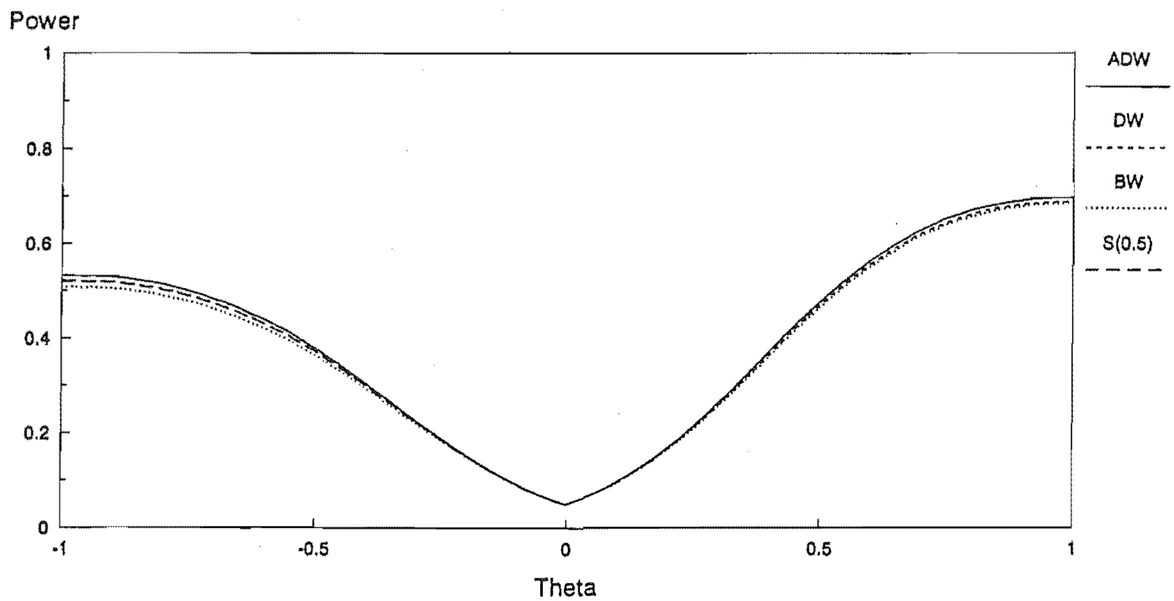


Figure 7.3.4
Power of DW Test against MA(1) Errors
Watson's Data; T=20; 5% Size

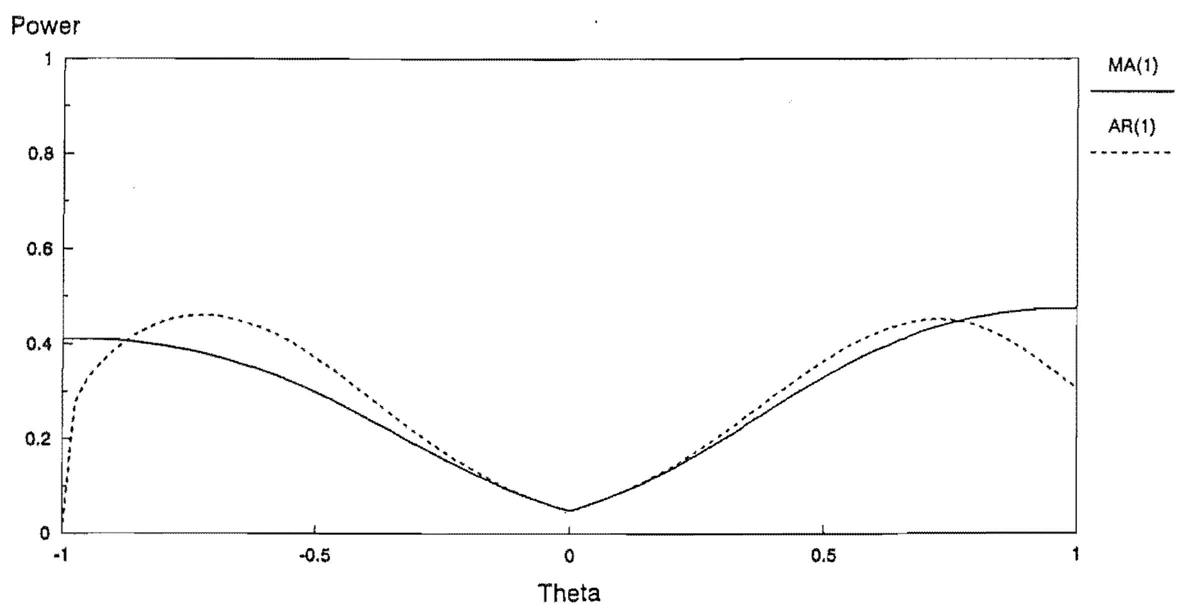


Figure 7.3.5
Power of S(0.5) Test against MA(1) Errors
Watson's Data; T=20; 5% Size

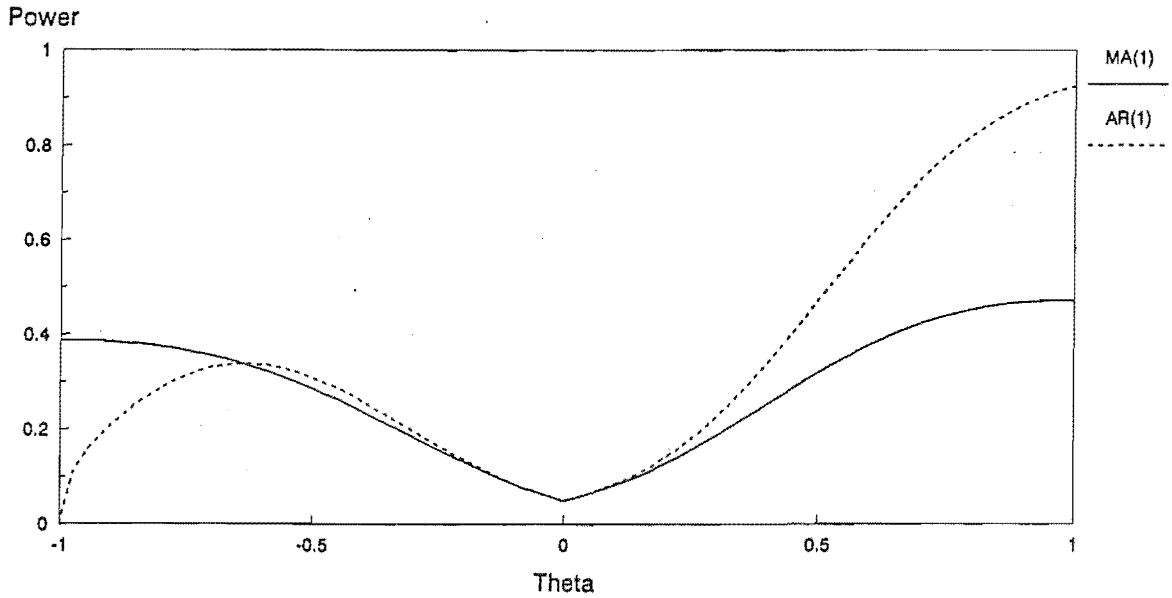
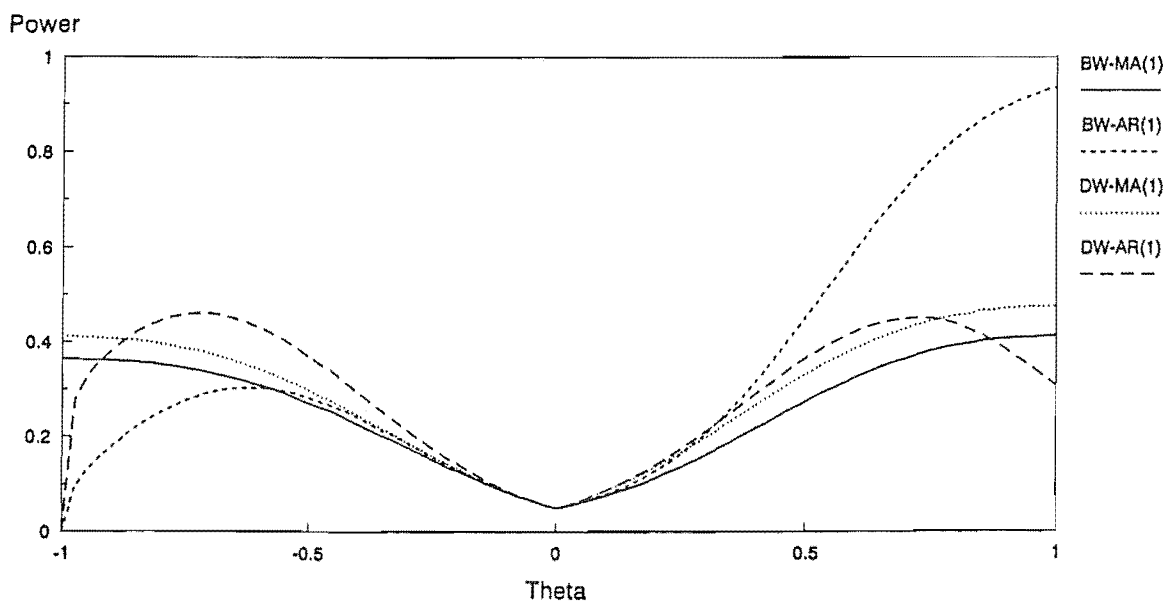


Figure 7.3.6
Power of BW and DW Tests against MA(1) Errors
Watson's Data; T=20; 5% Size



7.4 Testing With Seasonal Autoregressive Mis-specification.

Several authors have suggested that time series regressions using quarterly data could produce residual autocorrelation which has both first and fourth order components (see Harvey (1990, p.205), for example). This is entirely consistent with the standard rationale for the existence of a random disturbance term in a regression model. The possibility of simple AR(4) disturbances has been considered as a separate issue by Wallis (1972) and Vinod (1973) who proposed a fourth order generalisation of the Durbin-Watson (1950,1951) test, and by King (1984) who constructed the point optimal invariant test. Higher order processes have also been considered by Evans (1983) who studied tests against simple AR(j) errors for j=2,3,8 and 12, and by Durbin (1980) who used the special case of a regression on Fourier series to derive a sequential testing procedure for higher order processes. In addition, King (1989) presented a test designed to detect a simple AR(4) process when it is already known that AR(1) errors exist.

The aim of this Section is to take a step back from the analysis of King (1989) and seek the answers to two questions. First, how does the presence of a joint AR(1) and simple AR(4) error process affect the probability of detecting the AR(1) component? This will be answered by evaluating the power functions of several popular AR(1) tests under this form of mis-specification. The second question concerns the estimation efficiency of OLS relative to a feasible GLS estimator which might be used for final estimation, depending on the outcome of the AR(1) test. This issue could be addressed as a pre-testing problem by considering the risk, under some loss function, of the pre-test estimator and its components. The approach taken here, however, will focus on the spectral density of the error process.

7.4.1 Theoretical Issues

Consider the standard linear regression model of (1) with the following model for u:

$$(1-\phi_1 L)(1-\phi_4 L^4)u_t = \varepsilon_t \qquad t=1,2,\ldots,T \tag{10}$$

where L is the usual lag operator, such that $u_t(1-\phi_1 L) = u_t - \phi_1 u_{t-1}$ and $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$. Define $E(uu') = V$, where u is a $T\times 1$ vector with typical element u_t. Stationarity of (10) requires that $|\phi_1|$, $|\phi_4| < 1$ and these conditions will generally be imposed. This process can be seen as a restricted AR(5) scheme by writing (10) as

$$u_t = \phi_1 u_{t-1} + \phi_4 u_{t-4} - \phi_1\phi_4 u_{t-5} + \varepsilon_t \ . \tag{11}$$

To implement the procedure for evaluating test power which was outlined in chapter 6, the form of V is required. The covariance matrix used by King (1989) does not truly reflect (10) but the correct form can be derived from the Yule-Walker equations for this process. Denoting the autocovariance function by $\gamma_k = \gamma_{-k} = \text{cov}(u_t u_{t-k})$ gives

$$\gamma_0 = \phi_1\gamma_1 + \phi_4\gamma_4 - \phi_1\phi_4\gamma_5 + \sigma_\varepsilon^2$$

$$\gamma_1 = \phi_1\gamma_0 + \phi_4\gamma_3 - \phi_1\phi_4\gamma_4$$

$$\gamma_2 = \phi_1\gamma_1 + \phi_4\gamma_2 - \phi_1\phi_4\gamma_3$$

$$\gamma_3 = \phi_1\gamma_2 + \phi_4\gamma_1 - \phi_1\phi_4\gamma_2$$

$$\gamma_4 = \phi_1\gamma_3 + \phi_4\gamma_0 - \phi_1\phi_4\gamma_1$$

and $$\gamma_1 = \phi_1\gamma_{k-1} + \phi_4\gamma_{k-4} - \phi_1\phi_4\gamma_{k-5}\ ; \quad \text{for all } k > 4.$$

The simultaneous solution of these equations provides the autocovariance function and subsequent division by γ_0 gives the following autocorrelation function, where ρ_k represents the correlation between u_t and u_{t-k}:

$$\rho_0 = 1$$

$$\rho_1 = \phi_1(1+\phi_1^2\phi_4)/(1+\phi_1^4\phi_4)$$

$$\rho_2 = \phi_1^2(1+\phi_4)/(1+\phi_1^4\phi_4)$$

$$\rho_3 = \phi_1(\phi_1^2+\phi_4)/(1+\phi_1^4\phi_4)$$

$$\rho_4 = (\phi_1^4+\phi_4)/(1+\phi_1^4\phi_4)$$

$$\rho_k = \phi_1\rho_{k-1} + \phi_4\rho_{k-4} - \phi_1\phi_4\rho_{k-5}\ ; \qquad \text{for } k > 4$$

The scale factor was found by this method to be

$$\gamma_0 = \sigma_u^2 = \frac{\sigma_\varepsilon^2(1+\phi_1^4\phi_4)}{(1-\phi_1^2)(1+\phi_1^4\phi_4^3-\phi_1^4\phi_4-\phi_4^2)}.$$

It is immediately apparent that these expressions collapse to those for the well known AR(1) case when $\phi_4 = 0$.

By routinely testing data for the presence of unit roots, econometricians explicitly acknowledge the fact that many economic time series are non-stationary. Also widely accepted, is the virtual inevitability that relevant variables are omitted from many regression models. The clear implication of these two facts is that we may often encounter non-stationary residuals. To explore the power properties of these tests along the unit root boundary of the stationary parameter space we observe that, for this problem, each major section of this boundary depends on one of the autoregressive parameters only. In particular, the process in (10) becomes non-stationary when either ϕ_1 or ϕ_4 reaches unity in absolute value.

The power of each test was computed numerically along these line segments using a modification of the techniques suggested by Krämer and Zeisel (1990). When $\phi_1 = 1$, for example, $V = \iota\iota'$ where $\iota = (1,1,\ldots,1)'$ and MV=0 for regressions with an intercept. Thus all of the λ_j of (5) are zero and the power of the test is undefined. The limiting power as $\phi_1 \to 1$ can, however, be determined by replacing V with a transformation matrix W such that

$$W = \lim_{\phi_1 \to 1} (1-\phi_1)^{-1}(V-\iota\iota').$$

This matrix W can be shown to be a Toeplitz matrix with first column equal to

$$W_1 = \frac{\phi_4-1}{\phi_4+1}\begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 + 2\phi_4 \\ 6 + 4\phi_4 \\ 7 + 6\phi_4 \\ 8 + 8\phi_4 \\ 9 + 10\phi_4 + 2\phi_4^2 \\ \vdots \end{bmatrix}.$$

It can also be easily seen by inspection of the autocorrelation function that $V(\pm 1, -1) = I_T$, where the arguments of V are the values of ϕ_1, ϕ_4. This means that the power of each test at these points is equal to the true size of the test. It can further be shown that at all points on the $\phi_4 = -1$ boundary, except the endpoints, the power of each test is either zero or one. This result follows directly from the findings of Krämer (1985) and Small (1991).

The final limiting case of interest is that defined by $\phi_4 = 1$. From the general form of the autocorrelation function it can be seen that imposing the condition gives:

$$\rho_0 = 1$$
$$\rho_1 = \phi_1(1+\phi_1^2)/(1+\phi_1^4)$$
$$\rho_2 = 2\phi_1^2/(1+\phi_1^4)$$
$$\rho_3 = \rho_1$$
$$\rho_4 = 1.$$

This pattern repeats indefinitely so that the individual autocorrelations must take only one of three values. The rank of V, and the number of non-zero eigenvalues in (5), is therefore three.

The power of each test was computed under these conditions for the entire range of data outlined above. ϕ_1 took values ranging from zero to 0.9. In every case, each of the three non-zero eigenvalues were found to be positive so that the powers of the tests were always zero. Thus, in the presence of a seasonal unit root, the tests studied here will never reject the null model under a broad (but not exhaustive) range of data conditions.

7.4.2 Numerical Evaluations

Using a sample size of 20, a thorough investigation was conducted across all tests and design matrices along 20 lines in the parameter space. A further more limited, study used a sample size of 60. This latter work confirmed the findings of previous studies (*e.g.*, King (1985)) that a larger sample increases the power of each test and reduces the power differences between the tests. The Tables and Figures mentioned here are located at the end of section 7.4.

The following features were observed with all seven data sets and each test and are stated relative to power against pure AR(1) disturbances. First, the true sizes of the tests are decreased (increased) by the introduction of a positive (negative) fourth order component. This can be seen in Figures 7.4.1 and 7.4.1 which use X1 and are representative of all design matrices under study. The only exceptions to this were for S(0.75) and BW when using X6, where slight size increases were registered as $\phi_4 \to 1$. On average, sizes were 29.5% as $\phi_4 \to -1$ and 0.87% as $\phi_4 \to 1$.

Second, serious losses of power were found when ϕ_4 fell in the interval (0.4,1.0) for all $\phi_1 > 0$ (see Figures 7.4.2 and 7.4.3, for example). This is not unexpected in view of the size effect noted above when $\phi_4 > 0$. No size corrections were made to the power functions, since $\phi_4 \neq 0$ is assumed to be a mis-specification. Table 7.4.3 provides power values which show that when $\phi_1 = 0.4$, the introduction of a fourth order component with $\phi_4 = 0.4$ reduces power from around 40% to 25% for the X4 matrix, while Tables 7.2.1 and 7.2.2 show that the reduction is even more severe for the X2 and X5 design matrices. Increasing ϕ_4 to 0.6 further reduces power to around 15% while when $\phi_4 = 0.8$ power was generally about 7%.

The third feature of the numerical results is that the power of all tests falls with increasingly negative ϕ_4, when this parameter falls in the interval (-1,-0.4), for all $\phi_1 > 0$. In this region the power reduction is somewhat less serious, being offset by increased size. Figure 7.4.5 illustrates the effect of a strongly negative ϕ_4, while Figure 7.4.6 shows how ϕ_4 affects test power when the AR(1) component is strong ($\phi_1 = 0.8$).

7.4.3 Conclusion

This section has shown that the presence of a seasonal component in the autocorrelation structure of the errors from a linear regression model is likely to significantly reduce the power of the most popular exact tests for serial independence against AR(1) alternatives. When the seasonal component is negative the tests have been shown to have sizes which are much larger than the nominal levels. A further effect of this mis-specification is a flattening of the power curve so that the power "advantage" of increased size only occurs at low values of the first order AR parameter. When the fourth order parameter is positive test size and power are severely reduced. There exists, therefore, a strong possibility that strongly autocorrelated errors will not be detected by the tests considered in regressions which use quarterly data.

Table 7.4.1
Power of ADW and s(0.75) Tests with Seasonal Component; 5% Size
CPI Data (X2)

$\phi 4$	ADW	s(0.75)
$\phi 1=0.4$		
-1.0	0.590	0.570
-0.8	0.580	0.570
-0.4	0.519	0.518
0.0	0.398	0.400
0.4	0.228	0.231
0.8	0.051	0.053
1.0	0.000	0.000
$\phi 1=0.8$		
-1.0	0.730	0.713
-0.8	0.803	0.804
-0.4	0.834	0.846
0.0	0.793	0.810
0.4	0.674	0.693
0.8	0.376	0.389
1.0	0.552	0.531

Table 7.4.2 Power of DW and s(0.5) Tests with Seasonal Component; 5% Size Lognormal Data (X5)		
$\phi 4$	DW	s(0.5)
$\phi 1=0.4$		
-1.0	0.586	0.576
-0.8	0.572	0.572
-0.4	0.503	0.516
0.0	0.381	0.392
0.4	0.219	0.217
0.8	0.051	0.045
1.0	0.000	0.000
$\phi 1=0.8$		
-1.0	0.726	0.718
-0.8	0.798	0.804
-0.4	0.835	0.850
0.0	0.798	0.814
0.4	0.681	0.693
0.8	0.368	0.361
1.0	0.560	0.546

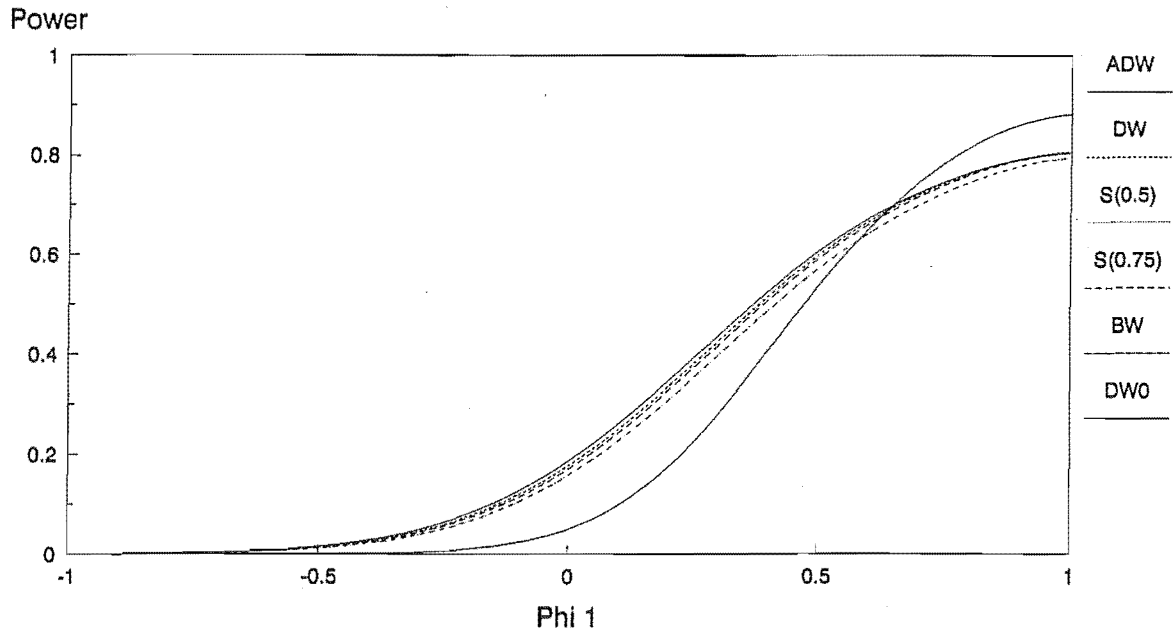
Table 7.4.3 Power of DW and BW Tests with Seasonal Component; 5% Size Uniform Data (X4)		
$\phi 4$	DW	BW
$\phi 1=0.4$		
-1.0	0.542	0.580
-0.8	0.524	0.563
-0.4	0.482	0.504
0.0	0.397	0.400
0.4	0.251	0.244
0.8	0.064	0.059
1.0	0.000	0.000
$\phi 1=0.8$		
-1.0	0.685	0.720
-0.8	0.782	0.815
-0.4	0.837	0.854
0.0	0.819	0.826
0.4	0.726	0.725
0.8	0.422	0.411
1.0	0.567	0.543

Table 7.4.4 Power of ADW and BW Tests with Seasonal Component; 5% Size Watson's Data (X6)		
$\phi 4$	DW	BW
$\phi 1=0.4$		
-1.0	0.549	0.414
-0.8	0.531	0.385
-0.4	0.444	0.348
0.0	0.299	0.317
0.4	0.132	0.265
0.8	0.018	0.153
1.0	0.000	0.001
$\phi 1=0.8$		
-1.0	0.695	0.570
-0.8	0.718	0.678
-0.4	0.631	0.791
0.0	0.450	0.828
0.4	0.238	0.805
0.8	0.059	0.627
1.0	0.253	0.000

Figure 7.4.1

Power Curves using Spirits Data; T = 20

5% Size; Phi4 = -0.8

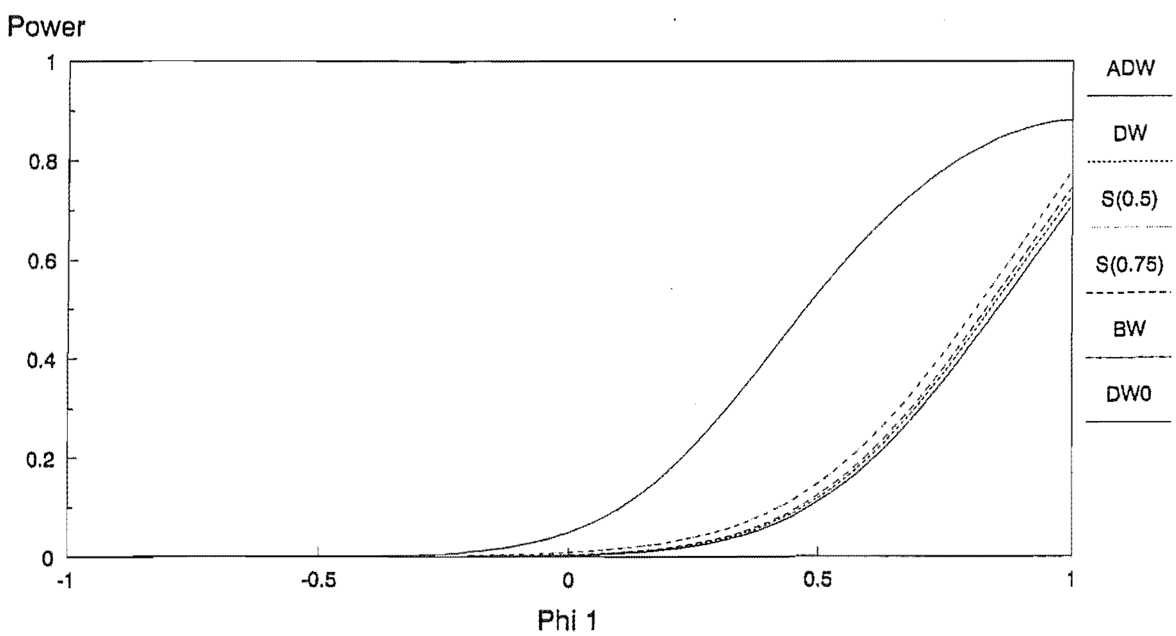


DW0 is DW Test When Phi4 = 0

Figure 7.4.2

Power Curves using Spirits Data; T = 20

5% Size; Phi4 = 0.8



DW0 is DW Test When Phi4 = 0

Figure 7.4.3
Power Curves using Unemployed Data ; T=20
5% Size; Phi4 = 0.6

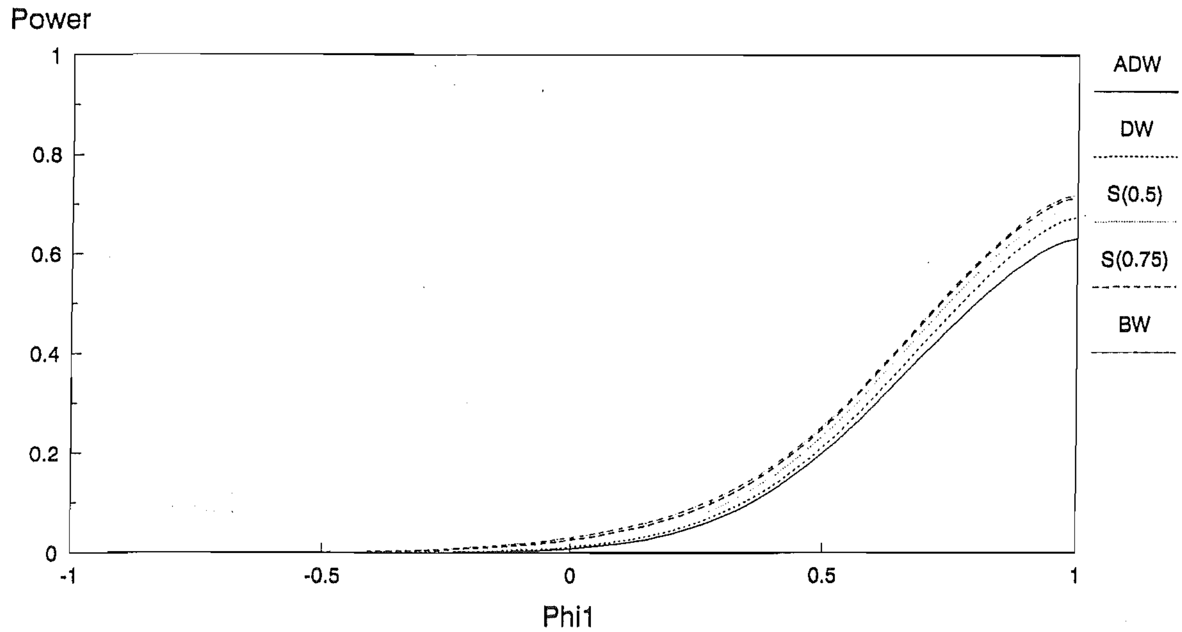


Figure 7.4.4
Power Curves using Unemployed Data ; T=20
5% Size; Phi4 = 0.8

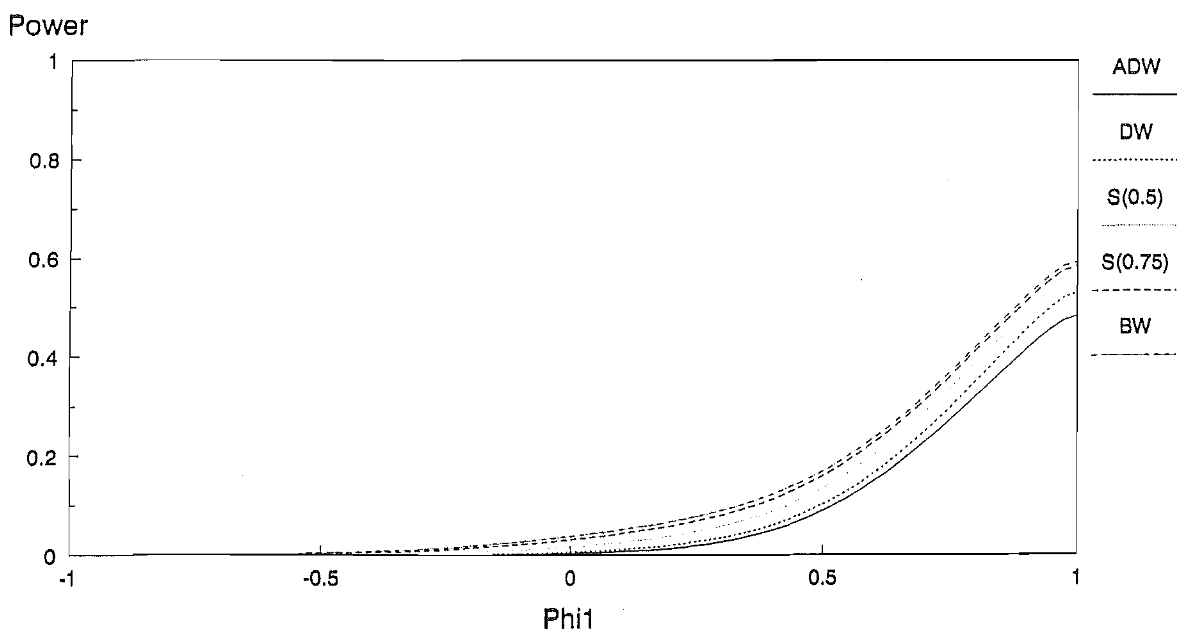


Figure 7.4.5
Power Curves using Normal Data ; T=20
5% Size; Phi4 = -1

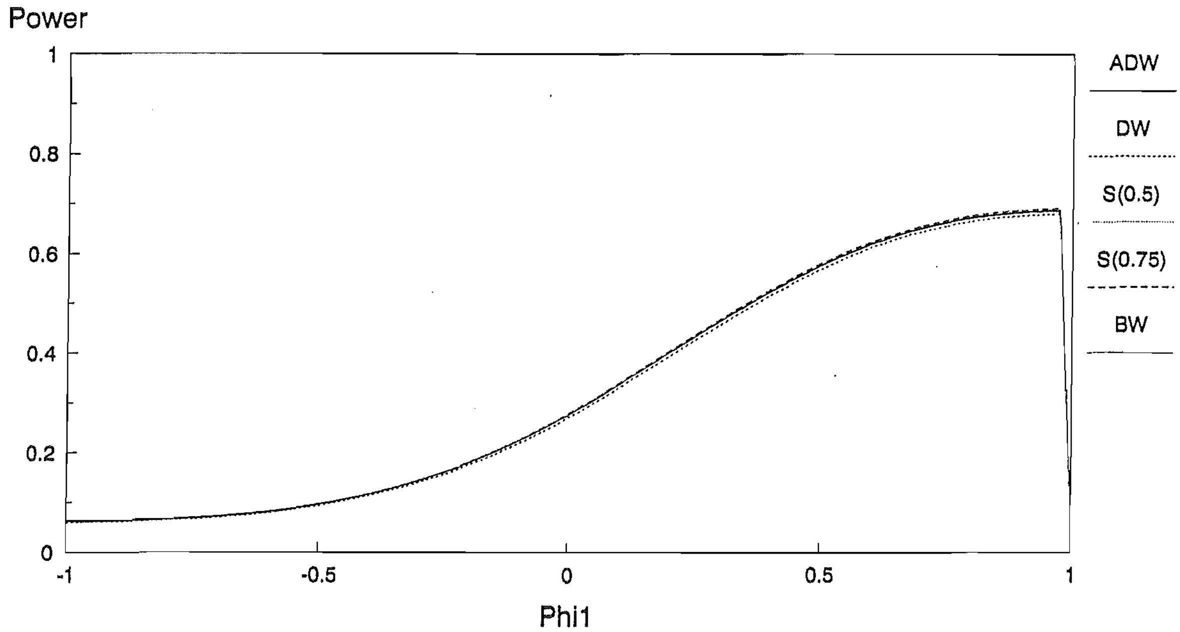
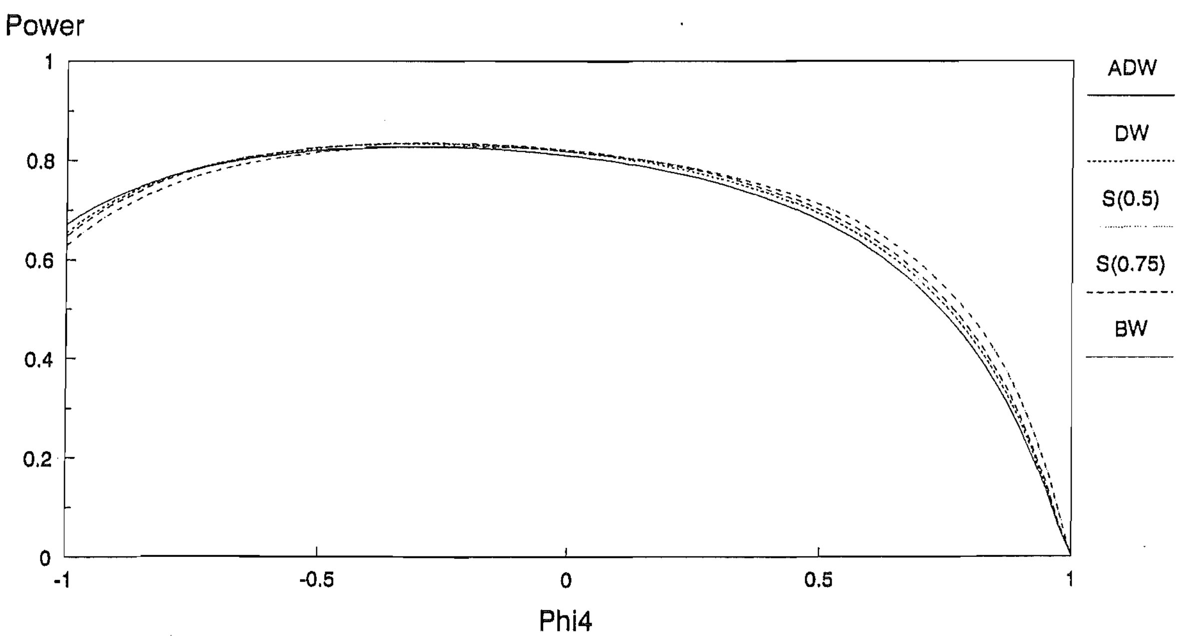


Figure 7.4.6
Power Curves using Spirits Data ; T=20
5% Size; Phi1 = 0.8



7.5 Conclusion

This chapter has considered the problem of detecting first order serial correlation when other error components are also present. It has been shown that, under each of the mis-specifications considered, the sizes and powers of several popular tests for AR(1) errors can be very different from those pertaining under the ideal (and unrealistic) conditions which describe a correctly specified model. This general finding has several important practical implications for applied researchers which we shall outline with the use of examples in which the result of using one of the tests considered in this chapter might lead an applied worker to mistakenly conclude that the regression errors were not serially correlated.

Let us consider the case in which a standard test against AR(1) errors fails to reject the null hypothesis of serial independence at a nominal 5% significance level. Suppose further that the researcher concludes, on the basis of this test, that the regression errors are serially independent. The analysis of section 7.2 has shown that, under certain conditions[6] the true size of this test could be as high as 35% when the true error process is AR(2). Thus, in this case, the nominal size is a very unreliable guide to the true Type I error probability. Exactly

[6] These conditions relate to using a Kadiyala based test with very "choppy" data such as Watson's (X6) matrix.

the same size effects can also occur if an AR(1) error process is combined with negative seasonal autocorrelation, as is shown by the results of section 7.4. In each of these cases, however, serial correlation **is present**, but in a different form from that being tested against. In particular, the first order component of the error process is zero. By neglecting to study further the time series characteristics of the regression residuals, the researcher is making a serious error which will result in inefficient parameter estimates and "t statistics" (for example) which do not follow a Student-t distribution.

Other situations which could lead to the incorrect assumption of serially independent regression disturbances have also been described in this chapter. In section 7.3 we showed that the presence of MA(1) errors results in relatively modest powers for the tests studied here. The optimality properties established by King (1983) for the DW and ADW tests in this model[7] may lead a researcher to conclude that failure to reject the null hypothesis in a DW test implies that neither AR(1) nor MA(1) errors are present. We have shown that for most data matrices considered, rejection relative frequencies are considerably lower for MA(1) models compared with AR(1) models when the respective parameters are the same size and larger than about 0.5.

A final case which can give rise to the erroneous conclusion

---

[7]Recall that King (1983) showed the DW test to be approximately LBI and the ADW test to be truly LBI in this model.

that serial correlation of the regression errors is not a concern has been shown, in section 7.4, to arise when AR(1) errors coexist with positive seasonal autocorrelation. Under these conditions the probability of rejecting serial independence is very low unless the first order effect is particularly strong (see Figure 7.4.2, for example).

In all of these cases, the ability of applied researchers to discover the time series properties of regression residuals is seriously impaired by the convenient fiction that the regression model is correctly specified. The consequences of ignoring serial correlation in this context are well known and have been noted above. In this chapter we have shown that the application of an exact test against AR(1) errors, although necessary, is a totally insufficient means of assessing the correlation structure of regression residuals.

CHAPTER 8

TESTING THE CONDITIONAL MEAN UNDER CONDITIONAL VARIANCE MIS-SPECIFICATION

8.1 Introduction

In this chapter we study the effect of a particular form of heteroscedasticity on the sizes and powers of a group of tests for serial independence. Specifically, we assume that the errors from a regression exhibit Generalised Autoregressive Conditional Heteroscedasticity, or GARCH. This topic is of direct relevance to the large numbers of contributors to the empirical finance literature who use various forms of the GARCH model (Engle (1982), Bollerslev (1986)) to represent the evolving volatilities of time series. In most applications modelling of the conditional mean is undertaken prior to the specification of the conditional variance (see Chou (1988), Hsieh (1989) or Baillie and Bollerslev (1989) for example). While this is a natural order in which to approach the modelling task, it also raises questions about the properties of tests used to detect autocorrelation in the mean equation, when the errors follow a GARCH process.

The class of GARCH models (which includes the ARCH model) was introduced in chapter 3, where tests of homoscedasticity against GARCH alternatives were discussed. In chapter 4 we surveyed the previous work on detecting serial correlation in the presence of GARCH, mentioning in particular work by Diebold (1986) and Wooldridge (1991) who have proposed tests with sizes which are claimed to be robust to GARCH. The point was made that neither of these authors provided convincing empirical support for their procedures.

A major aim of the study reported here is to correct this omission. The dependence of the tests' properties upon sample size is detailed, as is the effect of GARCH on the powers of the proposed tests. The "robust" procedures are compared with more standard tests using a range of data and two nominal test sizes.

In chapter 5 two models were introduced which allow for the combined presence of serial correlation in both the mean and the variance of a time series. The entire set of empirical measurements described above was conducted twice, once using the Weiss (1984) parameterisation, and again in the context of the Bera, Higgins and Lee (1992) random coefficient model. As a consequence, we are able to reveal some differences in the effect of these two specifications.

The chapter proceeds in section 8.2 with a brief restatement of the two methods of combining serial correlation and ARCH. This is followed by a discussion of the tests used in the empirical comparisons. section 8.4 describes the details of a Monte Carlo investigation that we have conducted, including a description of the data used. The results of this experiment are analysed in section 8.5, which is followed by some concluding comments.

8.2 The Models

The analysis of this chapter is based on the residuals from a regression model, rather than an observed series. Accordingly, we specify the basic model as

(1) $$y_t = x_t'\beta + u_t ; \qquad t=1,...,T,$$

where y_t is a scalar, x_t is a k×1 vector, β is a conformable parameter vector and u_t is a random disturbance. In many applications of relevance to this chapter x_t comprises a constant and a single regressor and this is reflected in the experimental design reported in section 8.4.

We want to allow u_t to exhibit serial correlation both in the mean and in the conditional variance. To achieve this we can use the framework pioneered by Weiss (1984) which simply involves appending[1] a GARCH innovation term to a standard ARMA process. Restricting attention to AR processes, we can write this as

(2) $$u_t = \sum_{k=1}^{r} \rho_k \, u_{t-k} + \varepsilon_t$$

where $\varepsilon_t | \psi_{t-1} \sim N(0,h_t)$ and

(3) $$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}.$$

The model described by (2) and (3) will be referred to as the

[1] As noted in chapter 5, Weiss in fact only used ARCH processes. Presumably GARCH possibilities would have been considered, however, if they had existed in 1984.

Weiss model. Clearly no restrictions are implied by this model on the order of the autoregressive component of the error term. Similarly, the GARCH process component of u_t is completely unrestricted.

An alternative to the Weiss model, proposed by Bera, Higgins and Lee (1992) (henceforth BHL) was also discussed in chapter 5. In this model the conditional heteroscedasticity is generated by allowing the parameters of a standard autoregression to be random. The ARCH version of this model is written as (1) with the following specification for u_t:

$$u_t = \sum_{j=1}^{p} \phi_{jt} u_{t-j} + \varepsilon_t \tag{4}$$

$$\phi_{jt} = \phi_j + \eta_{jt} \tag{5}$$

where $\eta_t = (\eta_{1t}, \eta_{2t}, \ldots, \eta_{pt})$ and $\eta_t \sim N(0,\Sigma)$ for some positive definite matrix Σ. The covariances of current and previous errors in the conditional variance equation can affect the estimate of the variance, for any given period, through the off-diagonal elements of Σ. This feature allows leverage effects to enter risk prediction (*i.e.* the **sign** of the lagged u_t's affects the conditional variance in the manner suggested by Nelson (1991)), something which cannot occur in the Weiss model. If Σ is diagonal the linear ARCH model of Engle (1982) is obtained. For the purpose of comparability with the Weiss model, as well as simplicity in experimental design, the study reported here uses only diagonal forms of Σ.

8.3 The Tests

Applied researchers in the empirical finance literature typically use a range of tests to diagnose deficiencies in the specification of the conditional mean of regression errors. The use of the exact tests studied in chapter 6 is relatively rare in this literature and the more thorough papers bolster these with a direct examination of the autocorrelation and partial autocorrelation functions (see Diebold and Nerlove (1989) for example). The portmanteau tests of Box and Pierce (1970) and Ljung and Box (1978) are heavily used, as are standard Lagrange Multiplier (LM) tests for serial correlation. In the last four years papers which use Diebold's (1986) correction have appeared (*e.g.* Hsieh (1989)) but the "robust" LM procedure of Wooldridge (1991) does not seem to have been routinely adopted.

In an effort to keep the size of the overall investigation manageable we concentrate on testing for serial independence against the alternative of simple AR(4) errors. There are several reasons for this choice. First, the strength of the GARCH, in practice, is positively related to the observation frequency, so that quarterly data is a more realistic choice than annual data. Second, there are several reasons, particularly under GARCH, for concluding that the powers of the Box-Pierce (BP) and Ljung-Box (LB) tests are higher against an AR(4) process than against an AR(1) process (see Box and Pierce (1970) p.1513, Diebold (1986) p.326). Finally, it would seem sensible to use a process for which analogues of the Durbin-Watson (DW) test are able to be constructed.

The fourth-order analogue of the DW test (denoted DW_4) was derived by Wallis (1972). The test statistic is defined as

$$d_4 = \frac{u' A_4 \, u}{u' u}$$

where $A_4 = A_m \otimes I_4$, A_m is a tri-diagonal m×m matrix with all non-zero off diagonal entries being -1, ones in the north-west and south-east corners and twos for the remaining diagonal elements. The dimension of A_m is one quarter of the sample size and $\otimes$ denotes the Kronecker product. Exact critical values for the Wallis test can be computed in the manner described in chapter 6 for the standard DW test. Similarly, exact powers can be found using Davies' (1980) algorithm, for example.

The only other exact test used in the study is a fourth-order generalisation of the s(0.75) test of King (1985). The test statistic for this test is given by

$$s_4(0.75) = \frac{u' Q \, u}{u' u}$$

where $Q = \Sigma - \Sigma X\left(X'\Sigma X\right)^{-1}X'\Sigma$, and Σ is the inverse of the theoretical covariance matrix of a simple AR(4) process assuming that $\rho_4 = 0.75$.

We now discuss the asymptotically justified tests for serial independence which are most commonly used in GARCH applications. These are the Ljung-Box (LB), Box-Pierce (BP) and Lagrange Multiplier (LM) tests, as well as "robust" versions of them.

The LB and BP tests are each based on sums of squared sample autocorrelations, differing only by a scaling factor. For the BP test against AR(4) errors, the statistic

$$Q(\hat{r}) = n\sum_{k=1}^{4}\hat{r}_k^2$$

is asymptotically distributed as χ_4^2 under the null hypothesis that the first four autocorrelations are jointly zero. Here $\hat{r}_k$ is the k^{th} sample autocorrelation. Ljung and Box (1978) proposed a modification to the BP statistic which was intended to provide a closer approximation to a quantity related to sums of the **true** squared autocorrelations. They suggested that treating

$$Q_m(\hat{r}) = n(n+2)\sum_{k=1}^{r}(n-k)^{-1}\hat{r}_k^2$$

as χ_4^2 under the null hypothesis would provide a more powerful test. Both of these test statistics draw on the finding of Bartlett (1946) that the r^{th} sample autocorrelations of a white noise process is asymptotically normal with zero mean and variance of $(n-r)/(n^2+2n)$. This is not true of an ARCH process, however, in which the variance of the r^{th} sample autocorrelation is shown by Milhöj (1985) to be $(1/n)(1 + \gamma_r^2 / \sigma^4)$, where γ_r^2 is the r^{th} autocovariance for the squared process and σ^4 is the unconditional fourth moment.

Because $\gamma_r^2 / \sigma^4 > 0$, the approximation of the variance of $\hat{r}_k$ by $1/n$ (as is done by Box and Pierce in constructing $Q(\hat{r})$) will systematically underestimate $var(\hat{r}_k)$, even in large samples. Furthermore, the additional factor of $(n-k)/(n+2)$ which is taken into account by the LB statistic $Q_m(\hat{r})$ reduces the value of the assumed variance still further. We should therefore expect that the size of the LB test is more severely affected by ARCH processes than the BP test.

Diebold (1986) suggested estimating γ_r^2 / σ^4 and using the estimates to construct adjusted versions of both the BP and LB tests. The test statistics for these are respectively

$$Q^a(\hat{r}) = n \sum_{k=1}^{r} \left(\frac{\hat{\sigma}^4}{\hat{\sigma}^4 + \hat{\gamma}_k^2} \right) \hat{r}_k , \quad \text{and}$$

$$Q_m^a(\hat{r}) = n(n+2) \sum_{k=1}^{r} \left(\frac{\hat{\sigma}^4}{\hat{\sigma}^4 + \hat{\gamma}_k^2} \right) \hat{r}_k/(n-k) \quad .$$

Exact expressions for σ^4 and $\hat{\gamma}_k^2$ are available for some conditional variance specifications (see Milhöj (1985) for example) but in practice these terms must be estimated and this is the method which was used in the Monte Carlo experiment reported in the next section.

When ARCH is present but the regression disturbances are serially independent, Diebold (1986) claims[1] that each of these statistics is asymptotically distributed as χ_p^2. This leaves several questions open. First, how are the rejection probabilities affected by this adjustment when the null hypothesis is not true? Second, what is the effect on the true

---

[1] Some doubt exists about the validity of this claim. The individual adjusted squared autocorrelations are certainly distributed as χ_1^2 but their independence questionable. This is because the adjustment includes autocovariances of the squared process indicating serial dependence through higher moments of the autocorrelations.

size of the tests of allowing for ARCH processes in this way when no ARCH effect is present? Third, is the adjustment also valid for GARCH processes? The first two of these questions will be addressed in the empirical study described below; we turn our attention to the third question now.

Suppose that $\varepsilon_t = \eta_t\sqrt{h_t}$, where $\eta_t \sim N(0,1)$ and

$$h_t = \alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{p}\beta_j h_{t-j}$$

$$= \alpha_0 + \alpha(L)\varepsilon_t^2 + \beta(L)h_t$$

where $\alpha(L) = \alpha_1 L + \ldots + \alpha_q L^q$ and $\beta(L) = \beta_1 L + \ldots + \beta_p L^p$. Then

$$h_t = \frac{\alpha_0}{1-\beta(1)} + \frac{\alpha(L)}{1-\beta(L)}\varepsilon_t^2$$

$$= \alpha_0^* + \sum_{i=1}^{\infty}\delta_i\varepsilon_{t-i}^2 \qquad (6)$$

where $\alpha_0^* = \alpha_0/1-\sum_{j=1}^{p}\beta_j$ and δ_i is the coefficient of L^i in the expansion of $\alpha(L)/(1-\beta(L)$. Equation (6) shows that a GARCH process is directly equivalent to an infinite order ARCH process. Thus Milhöj's (1985) representation of the variances of the sample autocorrelations for an ARCH process also applies to GARCH models. Recalling that Milhöj's expression was employed in the Diebold (1986) standard error correction, we conclude that this procedure is similarly valid for GARCH models.

We now conclude this section with a discussion of the LM test for AR(p) errors and an adjusted version of this test proposed by Wooldridge (1991). The general form of the LM test was described in chapter 2 and will not be repeated here. For conformity with the work reported below, consider the following AR(4) scheme for the u_t of (1):

(7) $$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \phi_3 u_{t-3} + \phi_4 u_{t-4} + \varepsilon_t$$

where it is assumed that the eigenvalues of the associated determinental polynomial lie within the unit circle, so that the process is stationary.

Under the null hypothesis H_0: $\phi_1=\phi_2=\phi_3=\phi_4=0$, and assuming that all other classical assumptions are satisfied, the Best Linear Unbiased Estimator is OLS, which is also the Maximum Likelihood Estimator. The nR^2 (or Outer Product Gradient) form of the LM test statistic for this problem is n times the uncentered coefficient of determination from a regression of the residuals, $\hat{u}_t$, from OLS estimation of (1) on the X matrix, and the first four lags of $\hat{u}_t$. Under the null, this statistic is asymptotically χ^2 with 4 degrees of freedom.

Wooldridge (1991), observing that this test is invalid in a dynamic model with conditional heteroscedasticity, proposed a general methodology for constructing tests which have **sizes** which are asymptotically robust in such cases. To focus on the practical application of Wooldridge's ideas we assume that the variables contained in X **do not**[2] include all lagged values of y_t.

Define $\lambda_t = (\hat{u}_{t-1},\ldots,\hat{u}_{t-4})$ and $h_t = E_t\left[Var(y_t|x_t)\right]$. In this

[2] The majority of Wooldridge's paper is based on models with completely specified dynamics, in which x_t contains the entire past history of y_t.

notation the standard LM test discussed above uses nR^2 from the OLS of:

$$\hat{u}_t = (x_t, \lambda_t)'\gamma + \eta_t$$

where γ is a suitably dimensioned parameter vector and η_t is a random error term. Wooldridge suggests initially weighting λ_t and x_t by dividing through by $\sqrt{h_t}$ where h_t is one's prior belief about the conditional variance function. In this study the weighting procedure is omitted because we wish to compare Wooldridge's LM test (WLM) with the standard procedure on the equivalent basis of ignorance about the presence (and therefore the form) of conditional heteroscedasticity. The Wooldridge procedure involves the following steps:

(i) Extract the 4×1 vector of residuals, r_t, from the vector regression of λ_t on x_t.

(ii) Define $\xi_t \equiv u_t r_t$ and extract the 4×1 vector of residuals, v_t, from the vector autoregression of ξ_t on $\xi_{t-1}, \ldots, \xi_{t-G}$.

(iii) Treat nR^2 from the regression of ι on v_t as asymptotically χ^2_4 under the null hypothesis (where ι is a vector of ones).

The number of lags in the VAR of step (ii) is arbitrary and will clearly affect the power of the test. Wooldridge recommends the use of "one or two (times) the integer part of $\sqrt[4]{T}$ ". Throughout this study four lags were used in this vector autoregression, a figure which corresponds with Woodridge's prescription for the sample sizes used below.

8.4 Experimental Design

To study the effect of conditional variance mis-specification on the size and power of the group of tests outlined above, a Monte Carlo study was conducted. All the work described below was conducted using 2000 replications which was found to produce reliable size figures for the exact tests used[3] prior to the addition of conditional heteroscedasticity.

Five design matrices were used, each of which comprised a constant and one other regressor. The first of these matrices contained the first two vectors from Watson's (1955) matrix which was discussed[4] in chapter 6. This matrix will be referred to as X6, to conform with the designation in that chapter but it should be noted that the matrix dimensions in this study are $T\times2$ rather than the previous $T\times3$ case. The non-constant regressors for the other four design matrices were based on the AR(1) process

(8) $$x_t = \lambda x_{t-1} + \varepsilon_t \;; \qquad t = 1,\ldots,T \;; \qquad \varepsilon_t \sim N(0,1).$$

The matrices, denoted X8, X9, X10 and X11, were constructed using the λ values of 0, 0.8, 1.0 and 1.02 respectively. These regressors are the same as those used by Engle, Hendry and Trumble (1985) and Lee and King (1991) and were incorporated in larger matrices by Giles, Giles and Wong (1992).

The majority of the study was conducted with a sample size

[3] This was verified by using the exact techniques outlined in chapter 6.

[4] Recall that the first "Watson vector" is a constant.

of 60, using the following basic model:

(9) $$y_t = x_t'\beta + u_t$$

(10) $$u_t = \rho_4 u_{t-4} + \varepsilon_t; \quad t=1,\ldots,T; \quad \varepsilon_t \sim N(0,1).$$

The power of first and fourth order variants of the DW and s(0.75) tests were evaluated, along with those of the BP and LB tests, their Diebold (1986) adjusted versions (BPA and LBA) and the LM and WLM tests. In each case two nominal sizes were used, namely 1% and 5%. All of the asymptotic tests were conducted against the general AR(4) alternative of (7).

For each design matrix the power of each test was evaluated at ten values of ρ_4 in the range [0,0.9] thus establishing benchmark power functions in correctly specified models. Conditional heteroscedasticity was then introduced into the model using both the Weiss and BHL specifications which are discussed in Section 8.2 above. The conditional variance function

(11) $$h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

was used with α_0 set to $1-\alpha_1$ and the following parameter sets for (α_1, β_1)

ARCH Models: (0.2,0) (0.4,0) (0.6,0) (0.8,0);

GARCH Models: (0.2,0.2) (0.2,0.4) (0.2,0.6) (0.2,0.8).

These parameter sets allow for a range of GARCH models which include some important cases in which the unconditional fourth moment of the disturbances is not finite. For the ARCH(1) model $3\alpha_1^2$ must be less than unity for the existence[6] of the

[6] This condition, and the following one for the GARCH(1,1) case, assumes conditional normality.

unconditional fourth moment of u_t, a condition which is violated for $\alpha_1 \geq 0.577$. In the GARCH(1,1) case the existence condition is $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$ and the GARCH parameter sets used here include one pairing (0.2,0.8) which violates this condition.

Apart from the basic format outlined above, two other avenues were explored. First, using sample sizes of $(i+5)^2$ for $i=1,2,\ldots,8$ the power of the LM and WLM tests were measured in the context of the X11 matrix and various GARCH specifications. The results of this experiment are used below to estimate the dependence of the power of these tests on sample size and conditional variance specification. Second, using the same design matrix with a sample of 60 observations the Monte Carlo experiment was repeated using innovations drawn from a conditional Student-t distribution with 4 degrees of freedom. The purpose of this variation was to assess the dependence of the main results on the conditional normality which is employed elsewhere.

8.5 Experimental Results

The results of the simulation experiments described above can be used to address three major questions. First, how does the presence of GARCH affect the properties of a group of tests for serial independence? Second, what influence does the sample size have on these properties? And third, to what extent are these conclusions dependent on the assumption of conditional normality of the errors? These topics are discussed in subsections 1, 2 and

3 respectively, which also address several other subsidiary issues. The relative effects of the Weiss and BHL models for combining AR and GARCH processes are discussed, as are the costs and benefits of "robustifying" the BP, LB and LM tests. All tables and graphs referred to are located at the end of this chapter.

8.5.1 Basic Properties

In this subsection the effects of adding GARCH innovations to the regression model are analysed. Unless otherwise stated, all tables and graphs referred to below are based on regressions with 60 observations and tests which are conducted at the 5% nominal significance level.

We begin by considering the exact DW and s(0.75) tests. Table 8.1 gives selected power values for these tests against AR(1) and AR(4) alternatives using X8 with 60 observations. The powers of the AR(1) tests are very low and generally decline as ρ_4 increases. This is not unexpected and correlates with the findings of section 7.4. The true sizes of these tests increase with the value of the ARCH parameter, α_1. It is, of course, unreasonable to expect tests against AR(1) errors to perform well against combined AR(4) - GARCH processes, given the results of section 7.4 in which the innovations are homoscedastic. The fourth order analogues of these tests, however, are known to perform very well against error processes such as (10), with white noise innovations, so the addition of GARCH innovations effectively constitutes only a single mis-specification.

Table 8.2 shows that the power function of the DW_4 test is reasonably robust to the addition of ARCH(1) innovations. This is particularly true of the endpoints of the curve; mid-range power is slightly reduced by ARCH. Similar effects were found for the $s_4(0.75)$ test across different data sets and for both the Weiss and BHL models, as Figure 8.1 illustrates. This figure also shows that the non-existence of the fourth moment flattens the power curve and increases the true size of an exact AR test in an AR-GARCH model (Figure 8.1b) but has no significant effect in an AR-ARCH model.

Turning our attention to the BP and LB portmanteau tests, we can see from Table 8.3 that the above conjecture about the relative effects of ARCH on the sizes of these tests is correct. The true size of the LB test is distorted upwards by more than than the BP test size, for a given degree of ARCH. When $\alpha_1=0.8$, for example, the BP with a nominal 5% size has a true size of 17%, while the corresponding LB test rejects the null incorrectly on 19.7% of occasions (it should be noted, however, that the LB test was found to have a higher true size even before the introduction of ARCH). In Figure 8.2a the power curves of both the BP and LB tests are shown to be flatter under ARCH, but eventually converge to unity as ρ_4 increases. Figure 8.2b shows that these results also apply to AR-GARCH models and with other design matrices. The following general conclusion is supported by all of the cases considered in this study: the major effect of conditional variance mis-specification on the BP and LB tests is a significant increase in their true sizes. Power effects are negligible for very strong autocorrelation (*i.e.*, for $\rho_4 \geq 0.9$), but can be larger or smaller than the correctly specified power for moderate values of ρ_4.

In the light of the above finding, one might expect that a correction which ensures that the size of a BP or LB test is robust to GARCH (such as that of Diebold (1986)) would be a major advantage in the models studied here. The empirical results show, however, that although the Diebold adjusted tests have more reliable size they also have less desirable power properties. Table 8.4 gives power values for the LB and LBA tests under ARCH innovations with the X9 matrix, in which the non-constant regressor is a sample from a white noise distribution. This table shows that for any substantial degrees of ARCH ($\alpha_1 \geq 0.4$) the sizes of the adjusted tests are much closer to their nominal levels, relative to the standard (unadjusted tests). Also of interest in this table are the substantial power differences found between the Weiss and BHL models at moderate to large ρ_4 values (*ceteris paribus*), the causes of which are unknown.

Figure 8.3 shows that, although the power of the Diebold adjusted LB test is relatively invariant to the ARCH parameter, α_1, (Figure 8.3a) the true size (and power) of the LBA test is substantially reduced when the errors contain GARCH processes (Figure 8.3b), other things being equal. Notice that this reduction occurs before the fourth moment ceases to exist (recall that this occurs here only when α_1=0.2 and β_1=0.8), although the effect is clearly much stronger when this does occur. It should also be noted that Figure 8.3 only compares the effect of different models on the powers of the LBA test; no inter-test comparisons are shown.

For the practitioner, however, a more interesting comparison is shown in Figure 8.4 which graphs the powers of the BP and the LB tests with their adjusted versions under specific ARCH (Figure 8.4a) and GARCH (8.4b) models. These graphs clearly show that the cost (in power terms) of obtaining a reasonably robust size by using the Diebold adjustment can be very high in all but the extreme regions of the (ρ_4) parameter space. This conclusion is reached regardless of the data or the choice of Weiss or BHL models. Furthermore, all four versions of the portmanteau test are markedly inferior to the standard fourth order DW test[7], as is clearly evident from Figure 8.4. This reinforces the view of Geweke (1988) that "the properties of "Q" are terrible in almost all econometric work", and suggests that efforts to devise powerful tests against general AR(p) alternatives could be of major benefit to applied finance researchers.

We now consider the standard LM test and its "robust" counterpart, the WLM test. In section 8.3 we discussed the rationale behind the WLM test which is designed to have the correct size asymptotically in the presence of conditional

---

[7] The s_4(0.75) test performs just as well as the DW test in these models.

heteroscedasticity. From Table 8.5 it is clear that the nominal size of the standard LM test is an unreliable guide to the true probability of a Type I error, even without the added complication of ARCH processes. At both the 1% and 5% nominal significance levels the LM test over rejects the null model and this problem is exacerbated by increasingly severe ARCH. The true sizes of the "robust" WLM test, however, while being invariant to ARCH, are much lower than their nominal levels, as is clear from Table 8.5.

In the above discussion of Table 8.4, attention was drawn to lower powers of the LB and LBA tests in the BHL model compared with the Weiss model. Figure 8.5 shows that this is again apparent in the power of the LM test. The models used in Figures 8.5a and 8.5b are identical apart from the method of combining the AR and ARCH components. It is clear that under the Weiss formulation the major effect of ARCH is on the size of the LM test, which is substantially above its correctly specified level. For the BHL model of Figure 8.5a, however, the size effects of ARCH are insignificant, compared with the consequences for mid-range power (and compared to the distortion from the nominal level which is 1% in these Figures).

To directly compare the relative merits of the LM and WLM tests, Figure 8.6 shows the power function of each test in four different models. First, in Figure 8.6a, ARCH innovations are used. In this case the size of the LM test is increased by ARCH, while the main effect on the WLM test is a reduction in power for moderately large ρ_4. More striking, however, is the very substantial difference between the power curves of the LM and WLM tests for a given model specification. As noted above, and confirmed in this graph, the WLM test has a true size which is somewhat lower than its nominal level. This does not adequately account for the very modest slope of the test's power curve, however. A very similar story is told by Figure 8.6b with respect to the inter-test comparison. In this case, however, the effect on the LM test of stronger conditional heteroscedasticity is reversed, with larger β_1 values tending to **increase** the size of the test. For all design matrices used, the direction of size distortion of the LM test was found to be upwards in ARCH models and downwards in GARCH models.

A final question to be addressed in this section concerns the influence of the design matrix in the above results. Subject to the caveat that we have only considered models with two regressors, the results appear to be relatively independent of the data. The broad findings remain valid for all of the X matrices used, as can be seen by comparing Figure 8.4 (a and b) with Figures 8.9 to 8.12 inclusive. These graphs show that the same overall relationships exist between the exact and both the standard and adjusted portmanteau tests, irrespective of the data.

We conclude this section with a summary of the most important findings. First, the exact tests (DW and s(0.75)) appear to be robust to the presence of GARCH processes provided that the autoregressive component of the errors is correctly specified. Second, the size of the LB test is more severely affected by GARCH than size of the BP test. Both versions of the portmanteau test strongly over-reject the null model in the presence of GARCH. Third, the Diebold adjustment, although asymptotically valid in either ARCH or GARCH models, has greater difficulty with the latter. In either case, the violation of the existence condition for the fourth moment seriously weakens the effectiveness of the adjustment. Finally, the LM test, although very powerful against most values of ρ_4, has an unacceptably high true size. The "robust" WLM test, by contrast, has very poor power with the sample sizes used here.

8.5.2 Sample Size Effects

From the previous section it is clear that the LM and WLM tests perform substantially worse than the other tests considered in the presence of GARCH. For the LM test the true size is unacceptably high, while the WLM test has serious power problems. Both of these tests, however, are only asymptotically justified and it could therefore be argued that the major reason for this poor performance is an insufficient number of observations. To investigate this possibility a further study was conducted using data matrix X8.

The powers of the LM and WLM tests were evaluated using samples of $(i+5)^2$ observations where $i=1,2,...,8$ and including all of the ARCH-GARCH parameter sets described in section 8.4

above. Nominal sizes of 5% were used and the autoregressive parameter ρ_4 was restricted to the values 0, 0.2, 0.4, 0.6 and 0.8. These evaluations generated a set of 360 cases (nine GARCH specifications for each of five AR(4) models and eight sample sizes) for each test. The observed rejection relative frequency was then regressed on the associated parameter set to provide a summary of the effects of sample size and GARCH specification. To allow for different effects under the null and alternative models, two separate models were estimated for each test, one using rejection relative frequencies from models for which ρ_4=0, and the other using the balance of the available data. Much of the detail of the estimation of these response surfaces follows similar work by Engle, Hendry and Trumble (1985) who were interested in tests against ARCH, rather than the tests against AR(4) alternatives studied here.

A logistic transformation was applied to the empirical rejection frequencies to ensure that predictions from the response models are restricted to the unit interval. We define: L(P) = log(P/(1-P)), where P is the observed rejection relative frequency. This transformation does not limit the domain of the dependent variable in any way.

Cox (1970) showed that A(L(P)), where A^2=MP(1-P) and M is the number of trials used to generate P, is approximately symmetrically distributed around zero with a unit variance. To exploit this result, a weighted least squares regression was fitted using A as the scaling factor for all of the variables in the regression. The SHAZAM (1993) econometrics computer program was used to obtain the following estimates, where t statistics appear in parentheses below their corresponding parameter, LM refers to L(P) where P is the rejection relative frequency of the LM test, and WLM is similarly defined.

$$A(LM)|_{\rho_4=0} = \underset{(0.97)}{0.490\alpha A} + \underset{(0.08)}{0.035\beta A} - \underset{(-1.44)}{21.380 A/T}$$

N=72, $R^2=0.42$

$$A(LM)|_{\rho_4\neq0} = \underset{(1.95)}{1.955\rho_4 A} - \underset{(-1.35)}{0.386\alpha A} - \underset{(-0.18)}{0.045\beta A} - \underset{(-0.87)}{7.427 A/T}$$

N=288, $R^2=0.33$

$$A(WLM)|_{\rho_4=0} = \underset{(0.36)}{0.284\alpha A} - \underset{(-0.68)}{0.480\beta A} - \underset{(-1.67)}{38.781 A/T}$$

N=72, $R^2=0.25$

$$A(WLM)|_{\rho_4\neq0} = \underset{(1.42)}{0.469\rho_4 A} - \underset{(-1.22)}{0.515\alpha A} - \underset{(-2.01)}{0.663\beta A} - \underset{(-2.16)}{25.365 A/T}$$

N=288, $R^2=0.25$

The GARCH parameters appear here as α and β without subscripts. It should also be noted that the quoted R^2 values were formed separately as the squared correlations between the observed and predicted power values after the logistic transformation was "undone". The above results demonstrate the consistency of each test through the negative coefficient on the inverse of the sample size for the regressions with $\rho_4 \neq 0$. We can also observe the dramatic difference between the slope (in the ρ_4 direction) of the LM test power function (1.96) and that of the WLM test (0.47). This comparison quantifies the observed differences in power curve slope noted in section 8.5.1 above. Furthermore, the power of the WLM test is significantly reduced by the GARCH parameter, β, whereas no such effect is apparent for the standard LM test.

8.5.3 Relaxing Conditional Normality

To assess the dependence of the above findings on the assumption of conditional normality in the error term a limited investigation was conducted using the X8 and X11 matrices with the Weiss model. These data can be thought of as bounding design matrices described by (8). For this study the conditional distribution of the ε_t of (10) and (11) was assumed to be Student-t with 4 degrees of freedom.

The conclusions of the main study with respect to the exact AR(4) tests remain valid with conditionally t_4 errors. An example of this is shown in Figure 8.7a where the DW test can be seen to maintain its assigned significance level and suffer only very minor losses in power in the presence of a strong ARCH effect.

The BP test (which has slightly more reliable size under ARCH than the LB test) is seriously affected by the relaxation of conditional normality. Table 8.6 shows that the size distortion induced by GARCH is aggravated by the presence of conditionally

Student-t_4 innovations and that the powers of both the BP and BPA tests are lower under this distribution, despite having higher size, for $\rho_4 \geq 0.5$ and $\beta \geq 0.6$. The comparison between the LB and LBA tests which is depicted in Figure 8.7b reveals that for a given parameter set these tests have power curves with very similar shapes, the main difference being the size distortion exhibited by the unadjusted test. Comparing Figures 8.3b and 8.7b we can further conclude that the Diebold size adjustment is more successful when the underlying distribution has heavier tails, but that the power curve is less steeply sloped in this case.

The LM test, which has severe size problems even without ARCH, suffers very badly from the relaxation of the conditional normality assumption, as can be seen from Table 8.7. The size distortion when conditionally Student-t_4 errors contain GARCH components can exceed 600% (from a nominal 5% to a true 33.6%). The power functions for different degrees of ARCH converge at reasonably low values of ρ_4, however, (see Figure 8.8a) and are always steeper than those of the WLM test (Figure 8.8b).

The conclusion from this section is that the relaxation of the assumption of conditional normality exacerbates the size problems of the BP, LB and LM tests but does not change their qualities relative to the proposed "robust" versions. These adjusted tests, while generally achieving their aim of lowering true size, remain markedly inferior by the criterion of power function slope. The exact tests stand out as being the ideal choice under the distributional assumption adopted here. It should be explicitly acknowledged, however, that the relatively restrictive nature of the AR process used in this study provides almost ideal conditions for the exact tests used.

8.6 Conclusion

In this chapter we have substantially clarified several issues related to the specification of the conditional mean of a regression model when the errors are conditionally heteroscedastic. It has been shown that the well known exact tests for simple autoregressive processes are outstandingly robust to the presence of GARCH effects. This conclusion is, of course, subject to the usual assumption that the alternative model is indeed a simple AR process of the appropriate order. We have also shown that the very frequently used BP and LB tests have sizes which are substantially greater than their nominal levels when conditional heteroscedasticity is present but that despite this increased size they are still less powerful than the exact tests for virtually all degrees of autocorrelation. The LM test for AR(4) errors grossly over rejects the null model even without GARCH, which makes the problem worse.

The two existing methods for correcting size distortion in the BP, LB and LM tests, due to Diebold (1986) and Wooldridge (1991) are generally successful in their stated aims, although Wooldridge's procedure tends to over correct. The power curves of these "robust" tests are much less steep than those of the standard tests, however, which raises serious doubts about the advisability of their use.

Finally, but very importantly, these conclusions appear not to depend on the assumption of conditional normality, having been also found using the thicker tailed Student-t distribution with 4 degrees of freedom. It is, of course, possible that a skewed distribution may alter some of the conclusions. This is a matter which is worthy of some further research. The null distribution of the DW test has been studied under a variety of skewed distributions and appears reasonably robust, as detailed in section 4.1. Relatively little attention, however, has been paid to the LM and LB tests under these conditions. The combination of GARCH effects and a skewed conditional distribution may well occur in practice and it would be valuable to know something about the properties of these tests under such conditions.

TABLE 8.1
Power of $s_1(0.5)$ Test with ARCH Errors

	Weiss Model		BHL Model	
rho4	1% Size	5% Size	1% Size	5% Size
$\alpha1=0.0$				
0.0	0.012	0.049	0.012	0.051
0.3	0.011	0.047	0.011	0.050
0.5	0.015	0.049	0.014	0.046
0.7	0.022	0.052	0.018	0.049
0.9	0.027	0.053	0.023	0.047
$\alpha1=0.2$				
0.0	0.024	0.074	0.021	0.084
0.3	0.020	0.070	0.022	0.080
0.5	0.024	0.069	0.028	0.078
0.7	0.027	0.070	0.028	0.072
0.9	0.028	0.055	0.028	0.059
$\alpha1=0.4$				
0.0	0.036	0.090	0.042	0.107
0.3	0.036	0.092	0.037	0.098
0.5	0.037	0.091	0.037	0.096
0.7	0.036	0.085	0.041	0.090
0.9	0.031	0.063	0.055	0.105
$\alpha1=0.6$				
0.0	0.056	0.121	0.055	0.129
0.3	0.052	0.114	0.053	0.127
0.5	0.053	0.110	0.053	0.113
0.7	0.048	0.100	0.060	0.115
0.9	0.037	0.069	0.082	0.137
$\alpha1=0.8$				
0.0	0.075	0.147	0.080	0.151
0.3	0.068	0.134	0.074	0.139
0.5	0.068	0.128	0.069	0.135
0.7	0.062	0.111	0.078	0.131
0.9	0.044	0.082	0.112	0.177

TABLE 8.2
Power of DW_4 Test with ARCH Errors

	Weiss Model		BHL Model	
rho4	1% Size	5% Size	1% Size	5% Size
$\alpha 1=0.0$				
0.0	0.006	0.049	0.011	0.050
0.3	0.416	0.677	0.413	0.676
0.5	0.889	0.967	0.895	0.972
0.7	0.994	0.999	0.999	1.000
0.9	1.000	1.000	1.000	1.000
$\alpha 1=0.2$				
0.0	0.006	0.054	0.011	0.047
0.3	0.405	0.669	0.388	0.645
0.5	0.895	0.966	0.883	0.961
0.7	0.995	0.999	0.995	0.998
0.9	1.000	1.000	1.000	1.000
$\alpha 1=0.4$				
0.0	0.008	0.052	0.010	0.047
0.3	0.394	0.661	0.371	0.631
0.5	0.891	0.965	0.861	0.951
0.7	0.994	1.000	0.992	0.999
0.9	1.000	1.000	1.000	1.000
$\alpha 1=0.6$				
0.0	0.010	0.053	0.013	0.047
0.3	0.385	0.659	0.332	0.604
0.5	0.890	0.963	0.824	0.934
0.7	0.993	1.000	0.985	0.997
0.9	1.000	1.000	1.000	1.000
$\alpha 1=0.8$				
0.0	0.015	0.053	0.018	0.057
0.3	0.378	0.648	0.321	0.563
0.5	0.883	0.961	0.781	0.922
0.7	0.992	0.999	0.973	0.993
0.9	1.000	1.000	0.995	0.999

TABLE 8.3
Power of BP and LB Tests with ARCH Errors
Data Matrix X9; Weiss Model

rho4	BP(1)[1]	BP(5)	LB(1)	LB(5)
$\alpha 1=0.0$				
0.0	0.007	0.044	0.011	0.060
0.3	0.165	0.348	0.201	0.398
0.5	0.603	0.790	0.664	0.831
0.7	0.936	0.978	0.953	0.982
0.9	0.997	0.999	0.998	1.000
$\alpha 1=0.2$				
0.0	0.014	0.058	0.022	0.079
0.3	0.177	0.356	0.214	0.410
0.5	0.611	0.807	0.679	0.837
0.7	0.941	0.981	0.955	0.986
0.9	0.997	0.999	0.998	0.999
$\alpha 1=0.4$				
0.0	0.027	0.090	0.035	0.108
0.3	0.192	0.379	0.233	0.429
0.5	0.630	0.817	0.686	0.848
0.7	0.945	0.982	0.962	0.987
0.9	0.998	0.999	0.998	0.999
$\alpha 1=0.6$				
0.0	0.048	0.122	0.059	0.145
0.3	0.219	0.414	0.262	0.462
0.5	0.653	0.824	0.709	0.859
0.7	0.948	0.980	0.965	0.986
0.9	0.997	0.999	0.998	0.999
$\alpha 1=0.8$				
0.0	0.079	0.170	0.095	0.197
0.3	0.252	0.461	0.300	0.511
0.5	0.671	0.841	0.722	0.868
0.7	0.947	0.977	0.960	0.983
0.9	0.994	0.998	0.996	0.998

1. BP(1) refers to the BP test with a 1% nominal size. The other columns are similarly designated.

TABLE 8.4
Power of LB and LBA Tests with ARCH Errors
Data Matrix X9

	Weiss Model		BHL Model	
rho4	LB	LBA	LB	LBA
$\alpha1=0.0$				
0.0	0.060	0.069	0.059	0.070
0.3	0.398	0.215	0.401	0.221
0.5	0.831	0.558	0.834	0.564
0.7	0.982	0.854	0.987	0.861
0.9	1.000	0.959	1.000	0.964
$\alpha1=0.2$				
0.0	0.079	0.070	0.077	0.079
0.3	0.410	0.220	0.391	0.225
0.5	0.837	0.582	0.827	0.554
0.7	0.986	0.874	0.986	0.884
0.9	0.999	0.967	1.000	0.983
$\alpha1=0.4$				
0.0	0.108	0.073	0.110	0.076
0.3	0.429	0.223	0.409	0.230
0.5	0.848	0.593	0.837	0.546
0.7	0.987	0.899	0.978	0.871
0.9	0.999	0.971	0.997	0.974
$\alpha1=0.6$				
0.0	0.145	0.078	0.156	0.084
0.3	0.462	0.232	0.442	0.223
0.5	0.859	0.597	0.818	0.522
0.7	0.986	0.905	0.968	0.832
0.9	0.999	0.978	0.987	0.925
$\alpha1=0.8$				
0.0	0.197	0.083	0.211	0.102
0.3	0.511	0.237	0.463	0.228
0.5	0.868	0.602	0.805	0.492
0.7	0.983	0.906	0.945	0.760
0.9	0.998	0.978	0.956	0.826

TABLE 8.5
Power of LM and WLM Tests with ARCH Errors
Data Matrix X9; Weiss Model

rho4	LM(1)[1]	LM(5)	WLM(1)	WLM(5)
$\alpha 1=0.0$				
0.0	0.080	0.285	0.002	0.032
0.3	0.277	0.581	0.010	0.112
0.5	0.718	0.904	0.031	0.289
0.7	0.961	0.993	0.085	0.505
0.9	0.999	1.000	0.200	0.622
$\alpha 1=0.2$				
0.0	0.098	0.295	0.001	0.032
0.3	0.288	0.580	0.008	0.110
0.5	0.717	0.902	0.033	0.274
0.7	0.965	0.991	0.078	0.477
0.9	0.999	1.000	0.198	0.604
$\alpha 1=0.4$				
0.0	0.110	0.300	0.001	0.028
0.3	0.293	0.578	0.010	0.111
0.5	0.722	0.898	0.030	0.262
0.7	0.968	0.990	0.071	0.444
0.9	0.999	1.000	0.192	0.598
$\alpha 1=0.6$				
0.0	0.133	0.322	0.002	0.026
0.3	0.304	0.576	0.010	0.112
0.5	0.723	0.903	0.026	0.235
0.7	0.968	0.990	0.066	0.409
0.9	0.999	1.000	0.184	0.574
$\alpha 1=0.8$				
0.0	0.159	0.340	0.001	0.028
0.3	0.320	0.588	0.009	0.101
0.5	0.719	0.901	0.024	0.214
0.7	0.964	0.989	0.063	0.372
0.9	0.998	0.999	0.179	0.549

1. LM(1) refers to the LM test with a 1% nominal size. The other columns are similarly designated.

TABLE 8.6
Power of BP and BPA Tests under GARCH
Data Matrix X11; Weiss Model

	Normal Errors		Student t Errors	
Rho4	BP(5)[1]	BPA(5)	BP(5)	BPA(5)
$\beta 1=0.2$				
0.0	0.071	0.062	0.120	0.060
0.3	0.360	0.188	0.411	0.188
0.5	0.801	0.520	0.816	0.521
0.7	0.979	0.870	0.979	0.876
0.9	0.999	0.979	1.000	0.976
$\beta 1=0.4$				
0.0	0.078	0.050	0.152	0.048
0.3	0.370	0.169	0.427	0.196
0.5	0.796	0.502	0.814	0.513
0.7	0.979	0.874	0.976	0.866
0.9	0.999	0.982	0.999	0.977
$\beta 1=0.6$				
0.0	0.097	0.032	0.176	0.041
0.3	0.378	0.141	0.434	0.172
0.5	0.798	0.468	0.787	0.469
0.7	0.974	0.871	0.960	0.842
0.9	0.999	0.984	0.995	0.969
$\beta 1=0.8$				
0.0	0.139	0.017	0.255	0.055
0.3	0.382	0.099	0.444	0.140
0.5	0.760	0.393	0.706	0.354
0.7	0.954	0.792	0.891	0.648
0.9	0.995	0.964	0.960	0.863

1. BP(5) refers to the BP test with a 5% nominal size. The other columns are similarly designated.

TABLE 8.7
Power of LM Tests under GARCH
Data Matrix X11; Weiss Model

	Normal Errors		Student t Errors	
Rho4	LM(1)[1]	LM(5)	LM(1)	LM(5)
$\beta1=0.2$				
0.0	0.075	0.264	0.112	0.300
0.3	0.251	0.527	0.281	0.552
0.5	0.692	0.890	0.688	0.886
0.7	0.962	0.990	0.963	0.991
0.9	0.998	1.000	0.998	1.000
$\beta1=0.4$				
0.0	0.057	0.222	0.102	0.284
0.3	0.218	0.494	0.268	0.519
0.5	0.660	0.875	0.658	0.855
0.7	0.958	0.988	0.950	0.984
0.9	0.997	1.000	0.997	1.000
$\beta1=0.6$				
0.0	0.040	0.176	0.089	0.239
0.3	0.185	0.438	0.238	0.473
0.5	0.612	0.840	0.605	0.812
0.7	0.948	0.983	0.914	0.969
0.9	0.995	1.000	0.993	0.998
$\beta1=0.8$				
0.0	0.051	0.162	0.180	0.336
0.3	0.174	0.383	0.292	0.490
0.5	0.541	0.764	0.554	0.752
0.7	0.889	0.963	0.832	0.919
0.9	0.989	0.995	0.955	0.976

1. LM(1) refers to the LM test with a 1% nominal size. The other columns are similarly designated.

Figure 8.1a
Power of s(0.75) Test with AR(4)-ARCH(1) Errors
Data Matrix X10; Weiss Model

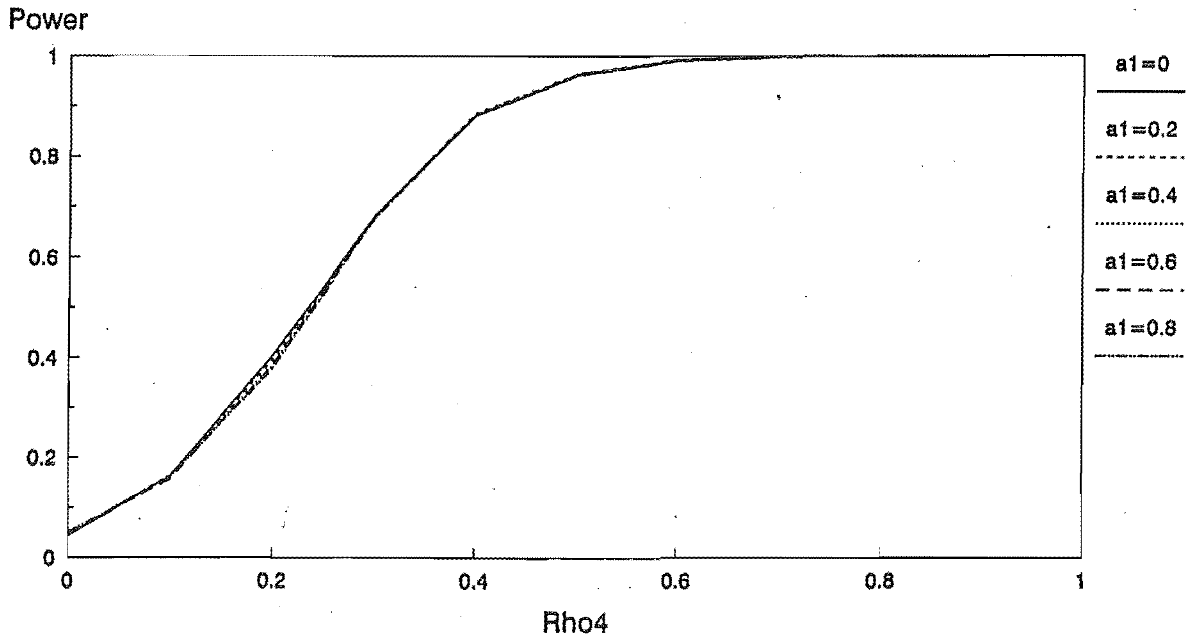


Figure 8.1b
Power of s(0.75) Test with AR(4)-GARCH(1,1) Errors
Data Matrix X10; BHL Model

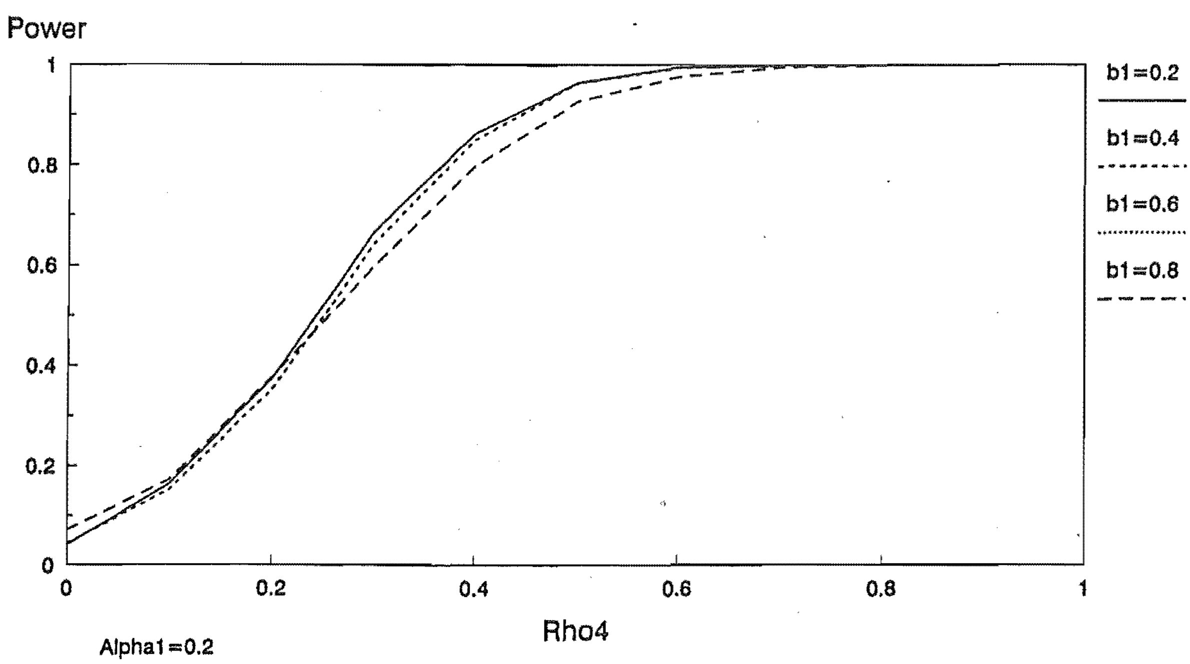


Figure 8.2a

Power of BP and LB Tests with AR(4)-ARCH(1) Errors
Data Matrix X9; Weiss Model

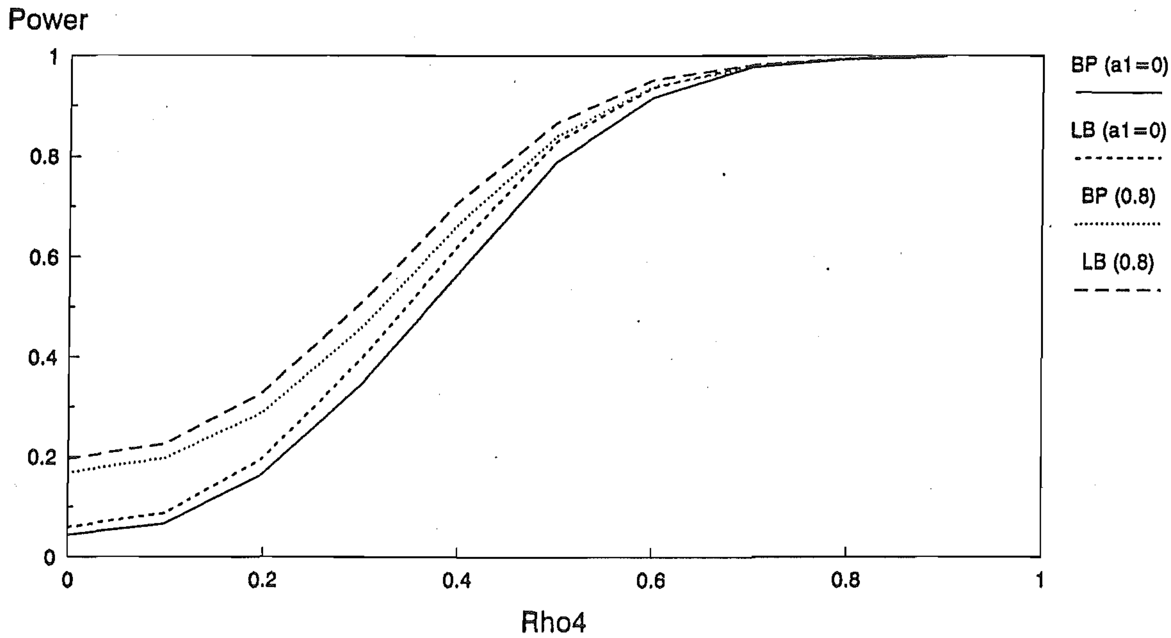


Figure 8.2b

Power of BP and LB Tests with AR(4)-GARCH(1,1) Errors
Data Matrix X11; BHL Model

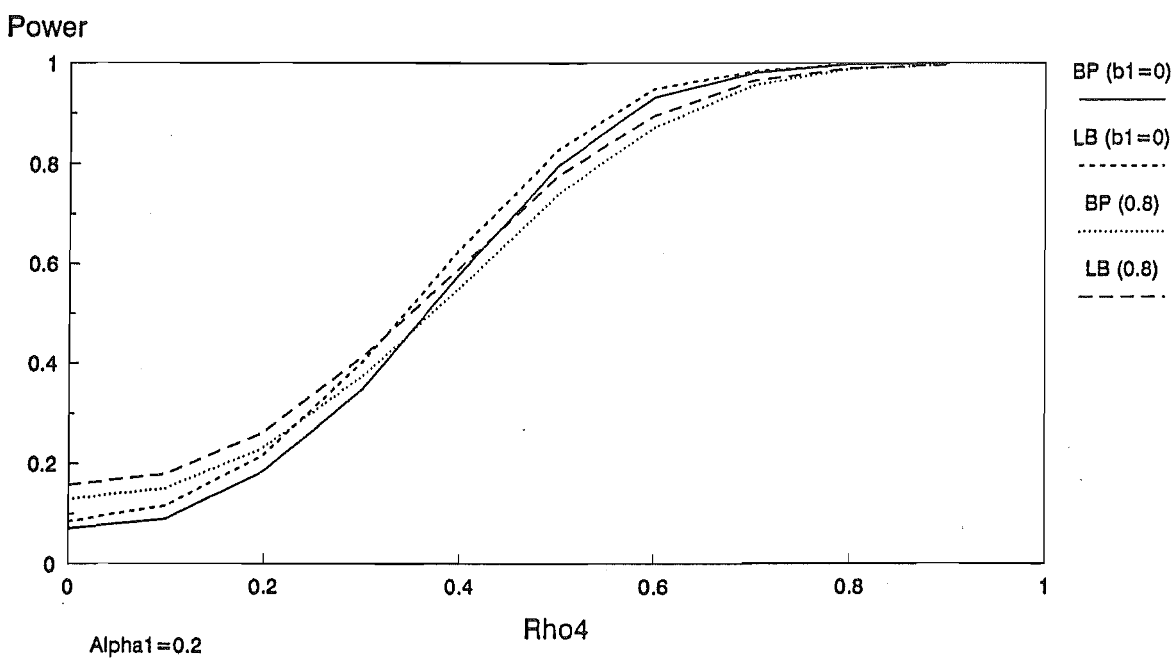


Figure 8.3a

Power of LBA Test with AR(4)-ARCH(1) Errors
Data Matrix X11; Weiss Model

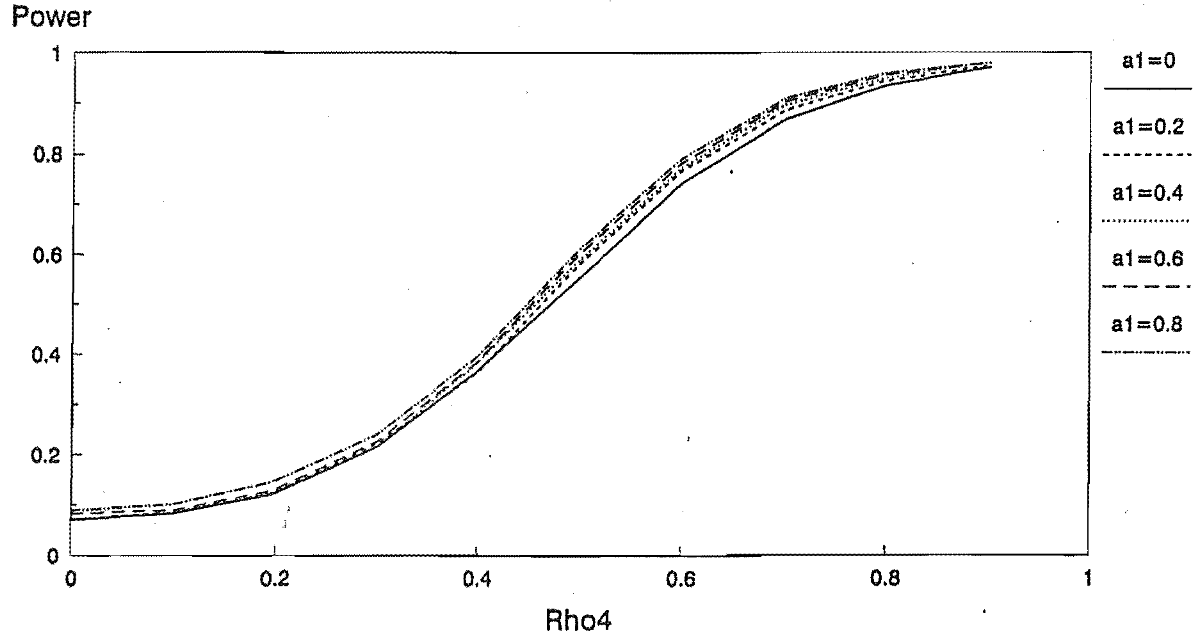


Figure 8.3b

Power of LBA Test with AR(4)-GARCH(1,1) Errors
Data Matrix X11; Weiss Model

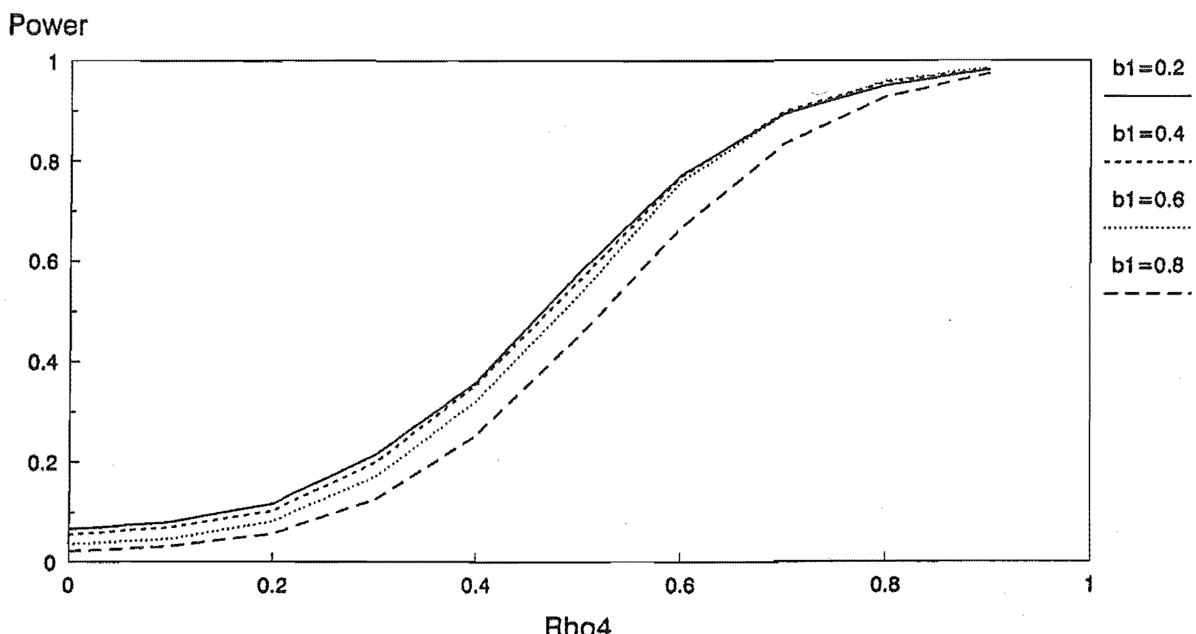


Figure 8.4a

Power of Several Tests with AR(4)-ARCH(1) Errors
Data Matrix X8; Weiss Model

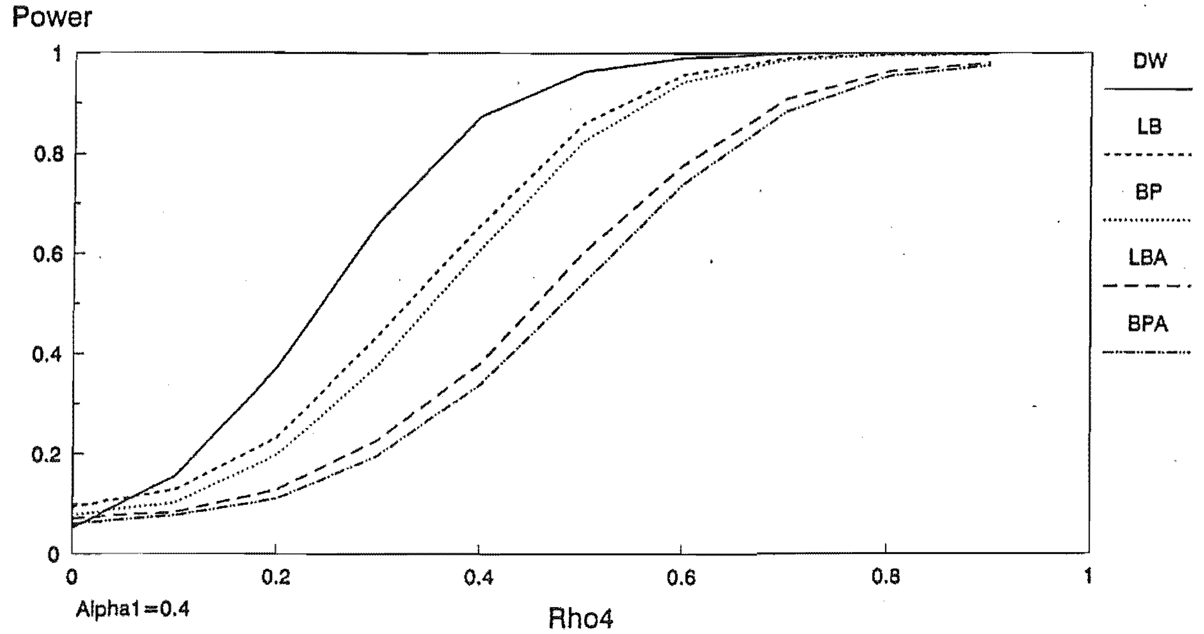


Figure 8.4b

Power of Several Tests with AR(4)-GARCH(1,1) Errors
Data Matrix X8; BHL Model

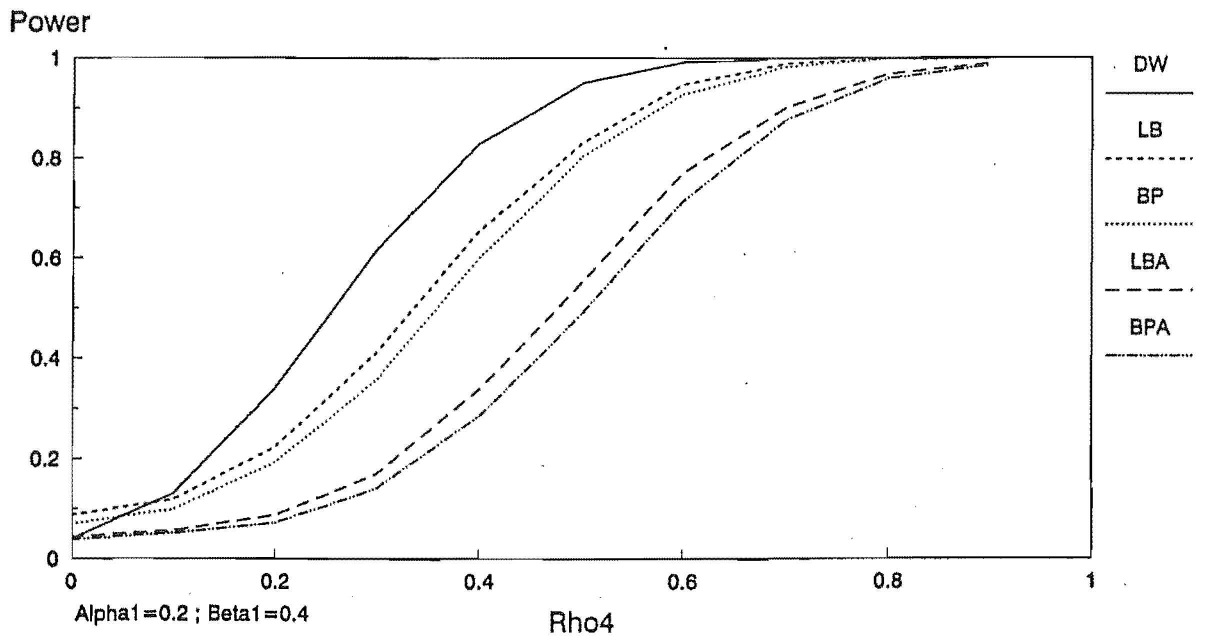


Figure 8.5a
Power of LM Test with AR(4)-ARCH(1) Errors
Data Matrix X8; BHL Model

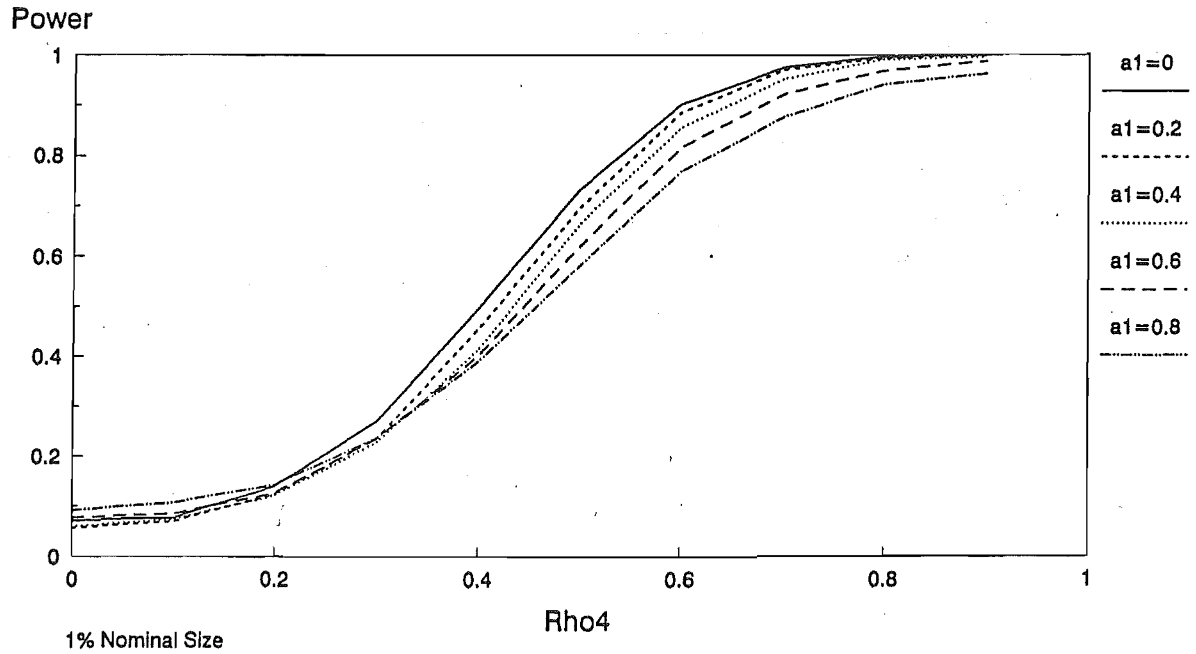


Figure 8.5b
Power of LM Test with AR(4)-ARCH(1) Errors
Data Matrix X8; Weiss Model

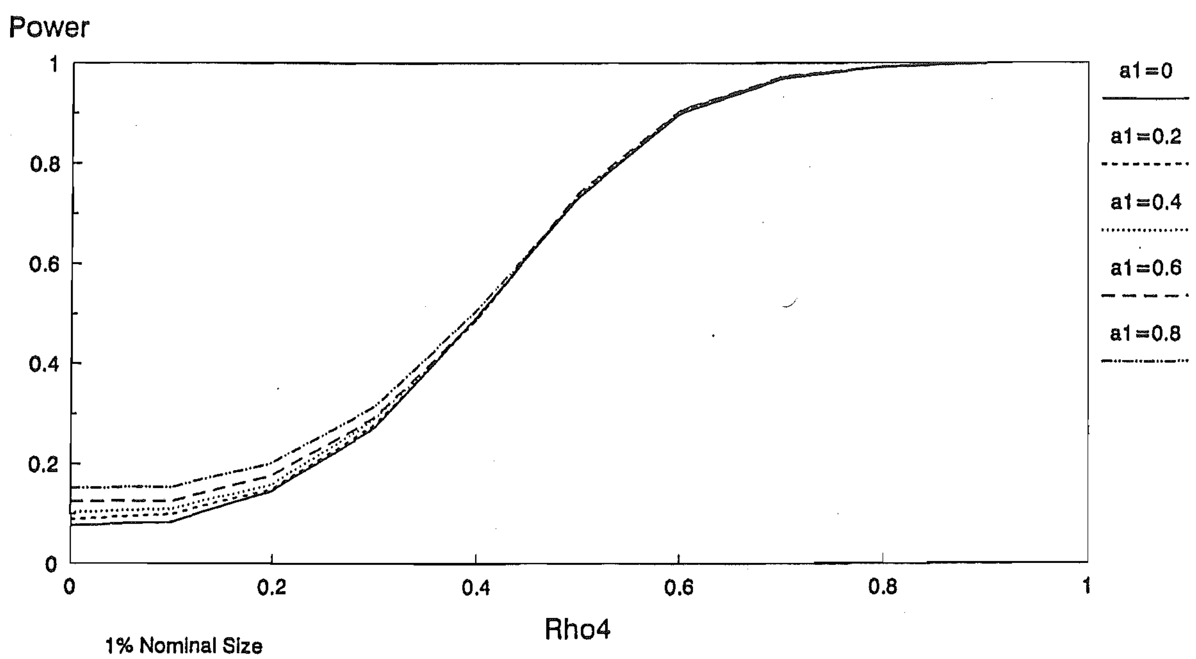
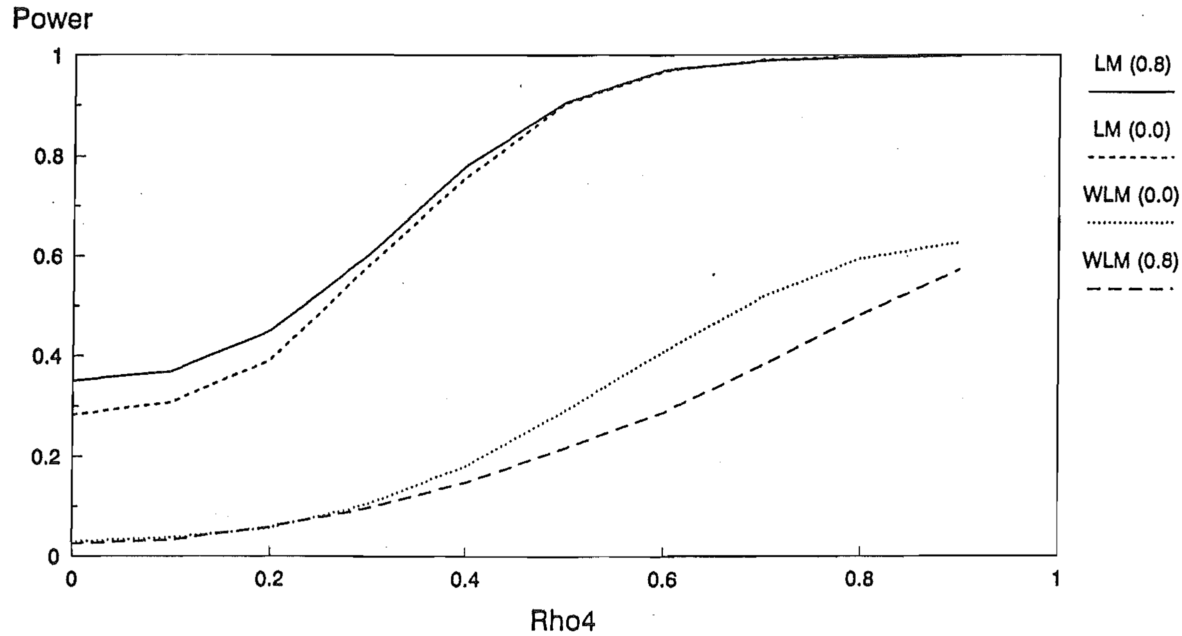


Figure 8.6a

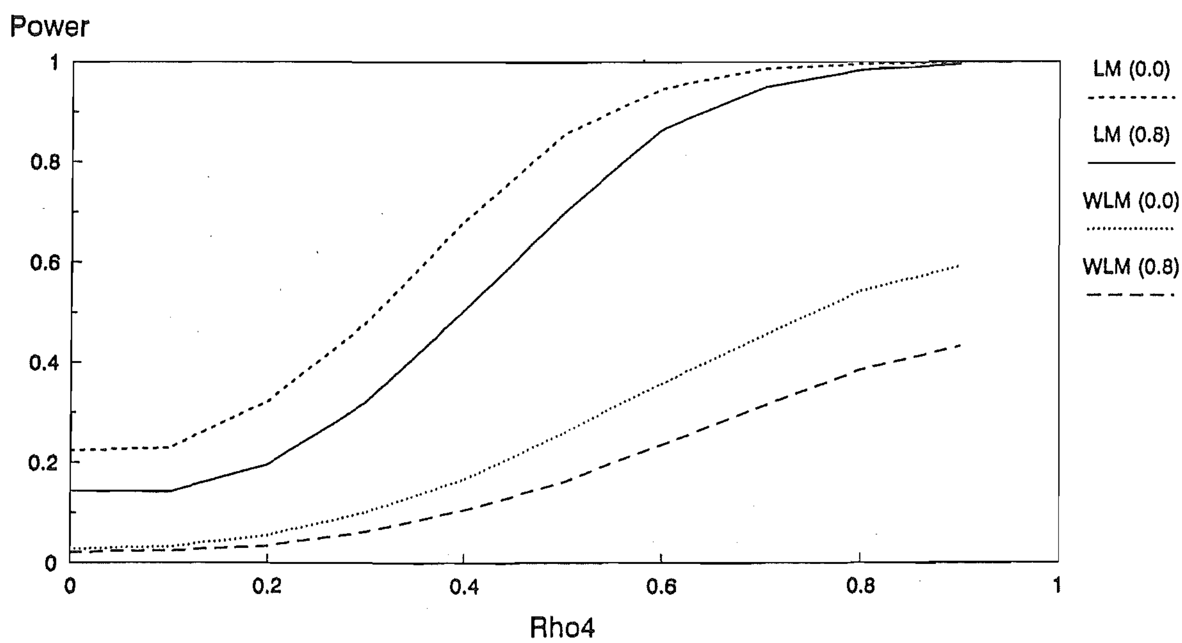
Power of LM and WLM Tests with AR(4)-ARCH(1) Errors
Data Matrix X10; Weiss Model



Bracketted Figure in LM (0.0) is ARCH Parameter

Figure 8.6b

Power of LM and WLM Tests with AR(4)-GARCH(1,1) Errors
Data Matrix X6; BHL Model



Bracketted Figure in LM (0.0) is GARCH Parameter

Figure 8.7a

Power of DW Test with Student t Errors and ARCH
Data Matrix X8; Weiss Model

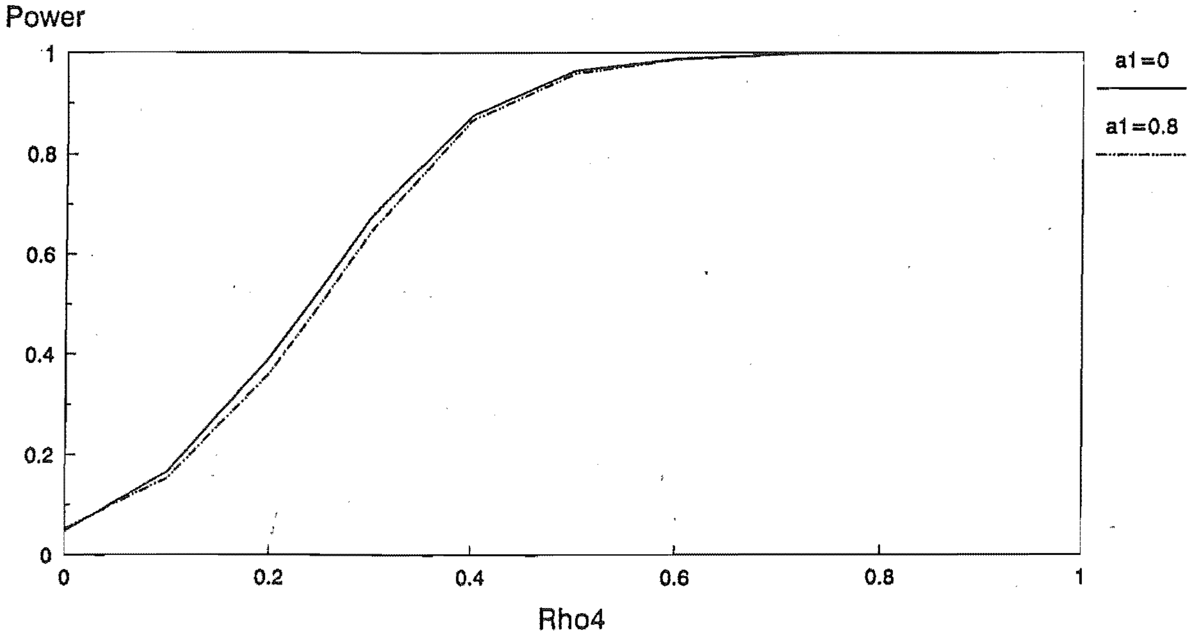
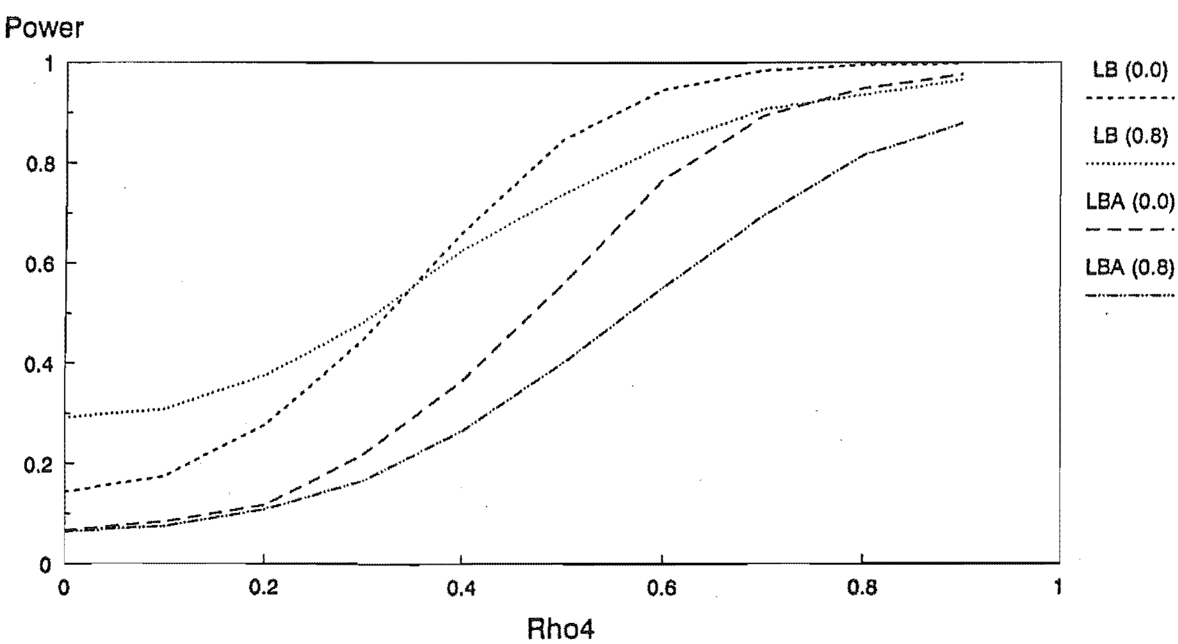


Figure 8.7b

Power of LB & LBA Tests with Student t Errors and GARCH
Data Matrix X11; Weiss Model



Bracketted Figure in LB (0.0) is GARCH Parameter

Figure 8.8a
Power of LM Test with Student t Errors and ARCH
Data Matrix X8; Weiss Model

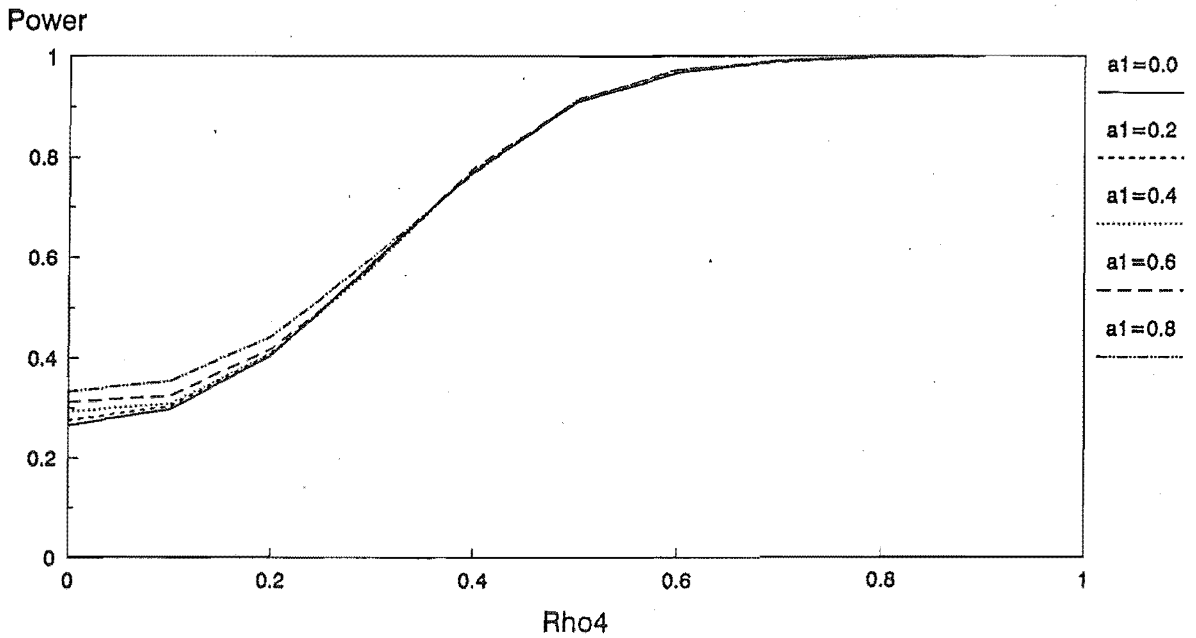
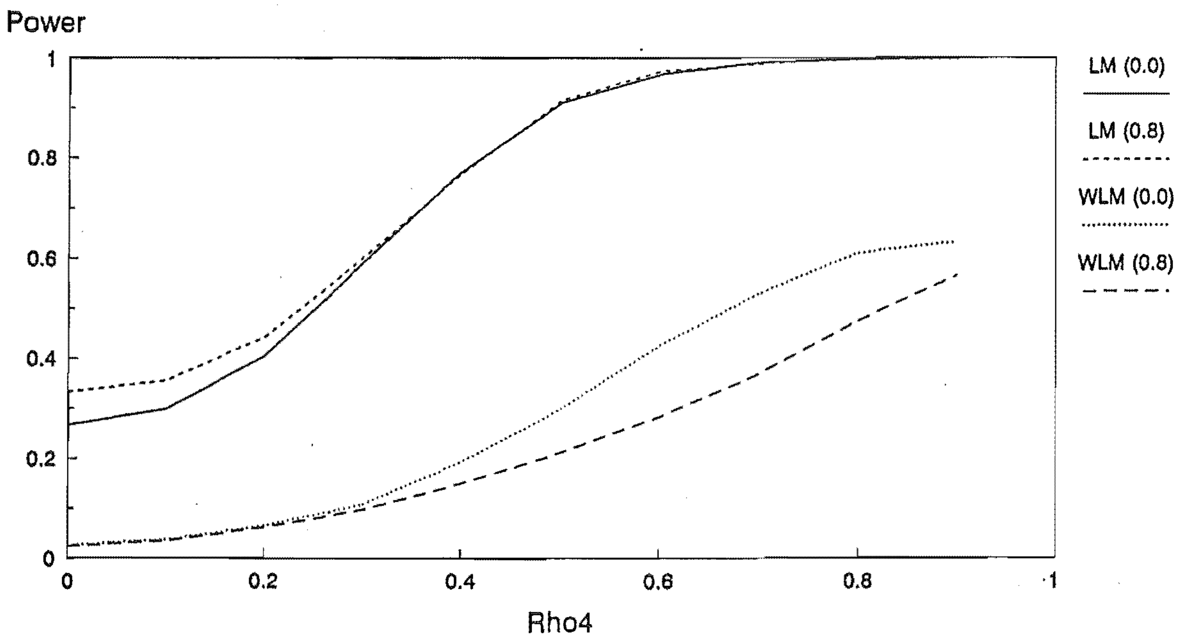


Figure 8.8b
Power of LM & WLM Tests with Student t Errors and ARCH
Data Matrix X8; Weiss Model



Bracketted Figure in LM (0.0) is ARCH Parameter

Figure 8.9a

Power of Several Tests with AR(4)-ARCH(1) Errors
Data Matrix X6; Weiss Model

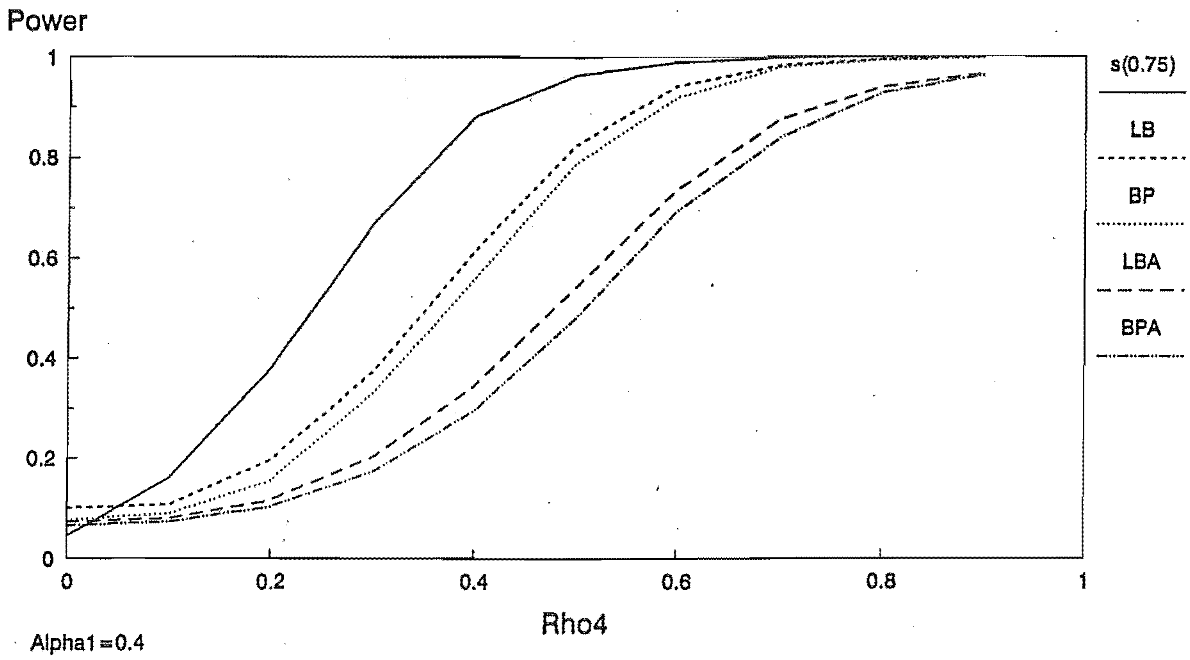


Figure 8.9b

Power of Several Tests with AR(4)-GARCH(1,1) Errors
Data Matrix X6; BHL Model

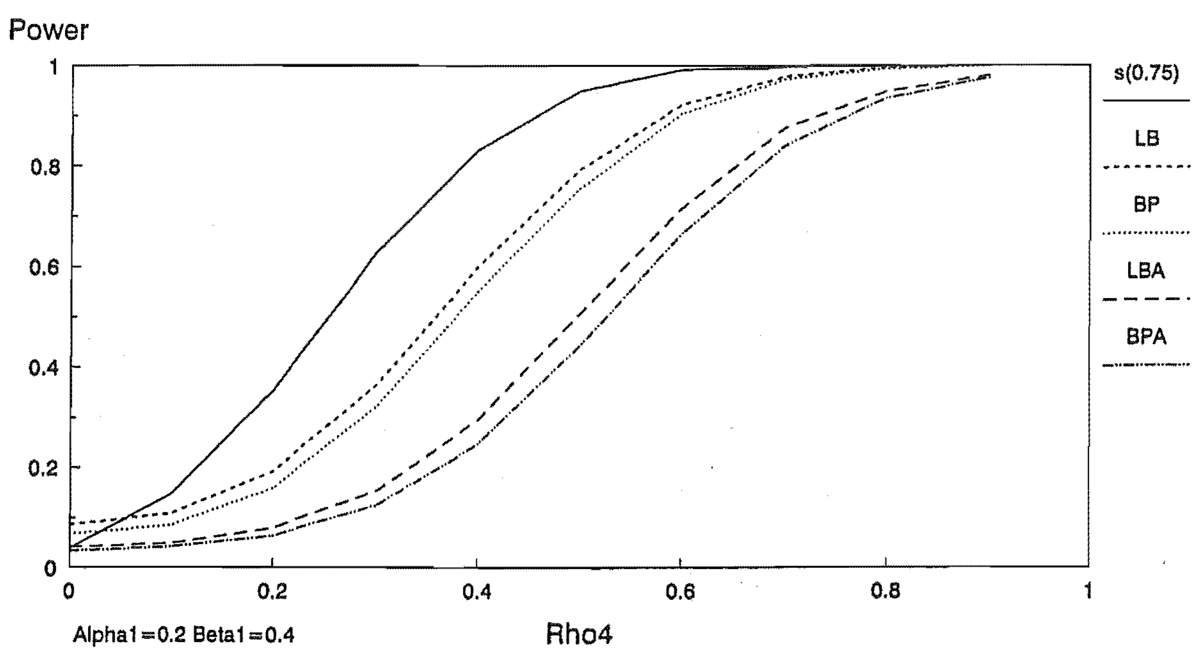


Figure 8.10a
Power of Several Tests with AR(4)-ARCH(1) Errors
Data Matrix X9; Weiss Model

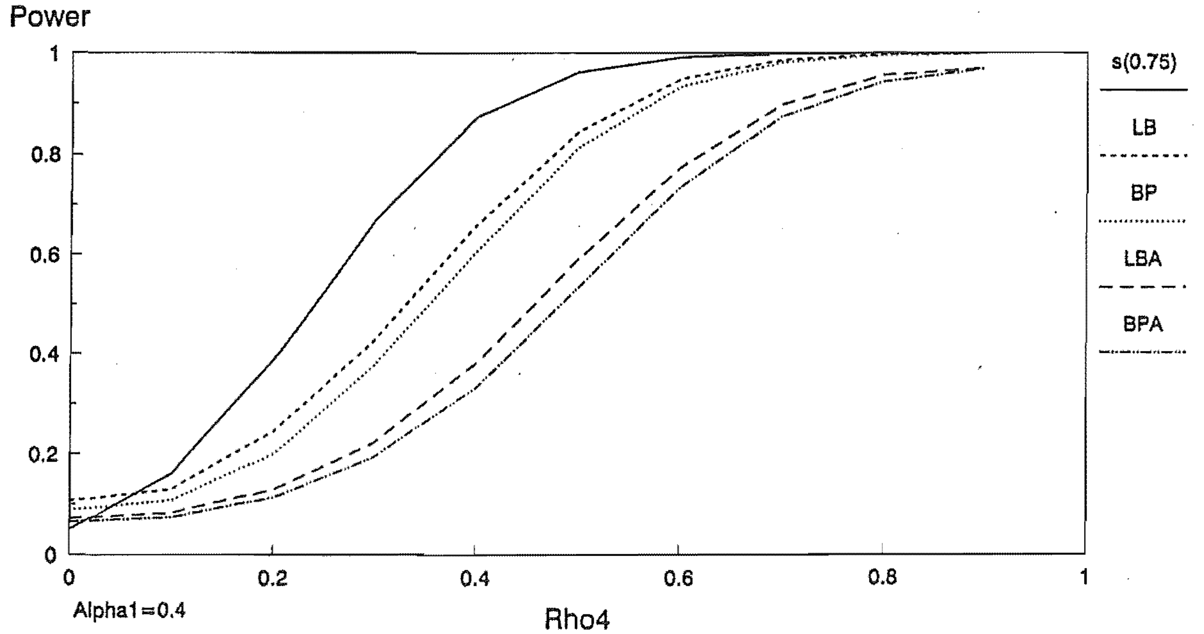


Figure 8.10b
Power of Several Tests with AR(4)-GARCH(1,1) Errors
Data Matrix X9; BHL Model

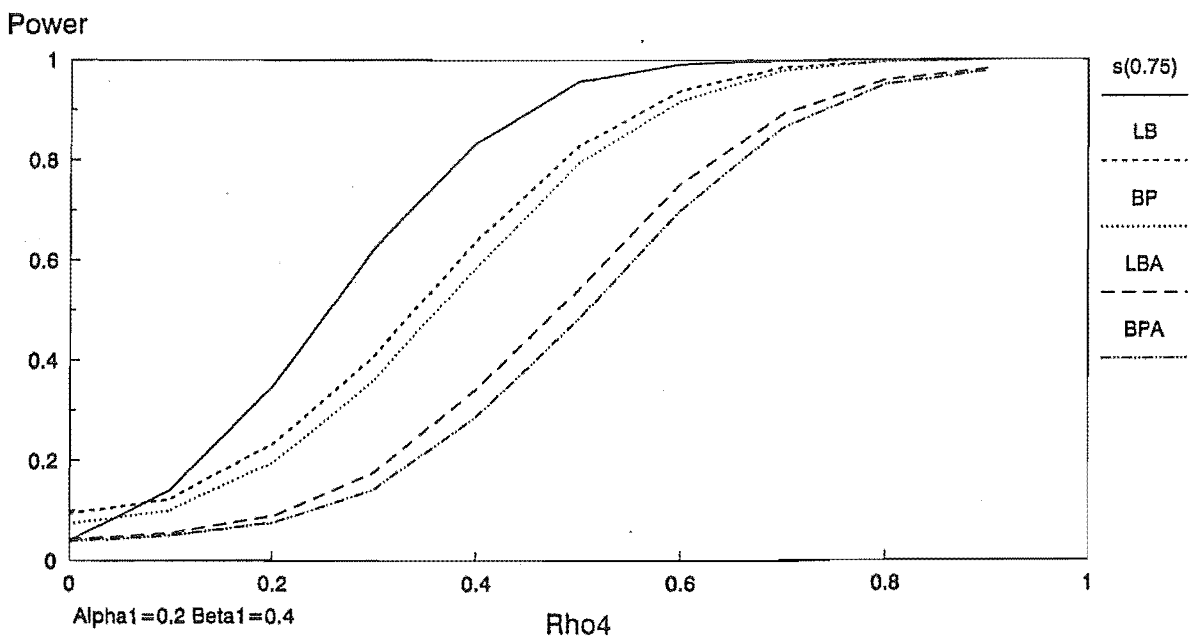


Figure 8.11a
Power of Several Tests with AR(4)-ARCH(1) Errors
Data Matrix X10; Weiss Model

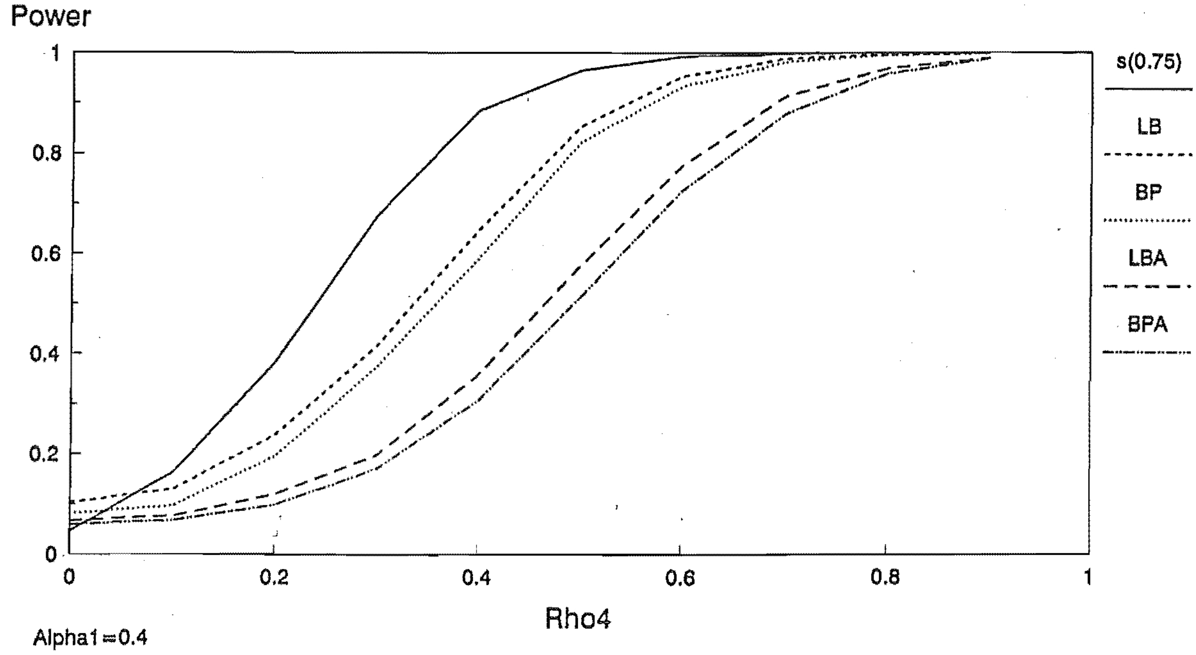


Figure 8.11b
Power of Several Tests with AR(4)-GARCH(1,1) Errors
Data Matrix X10; BHL Model

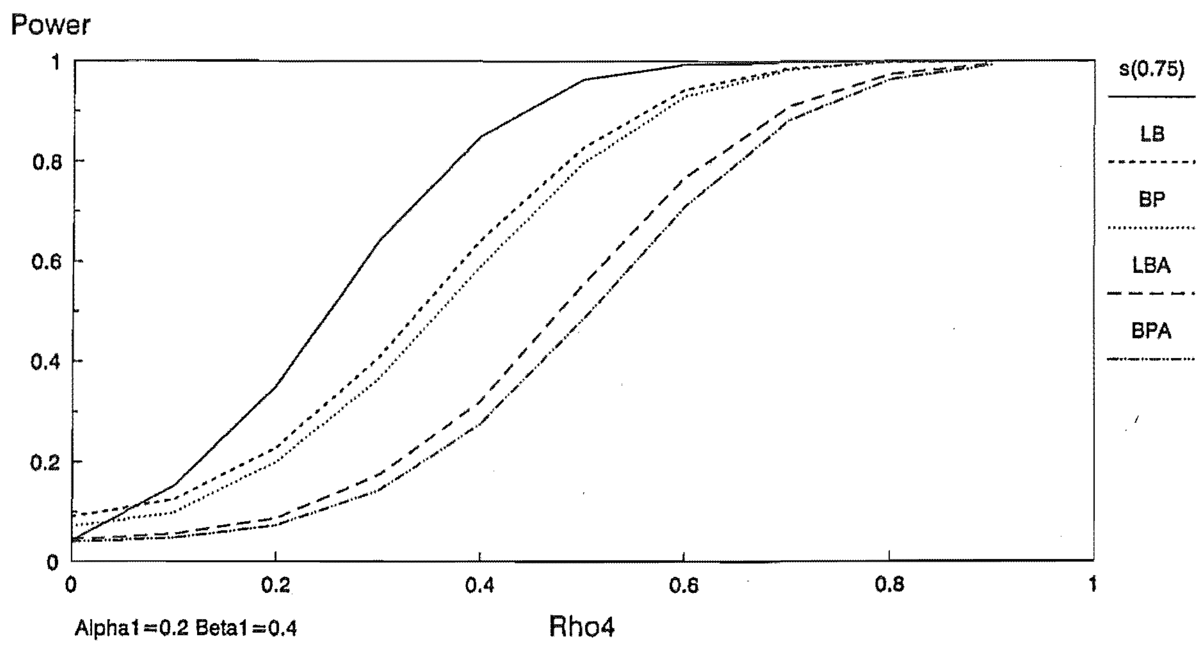
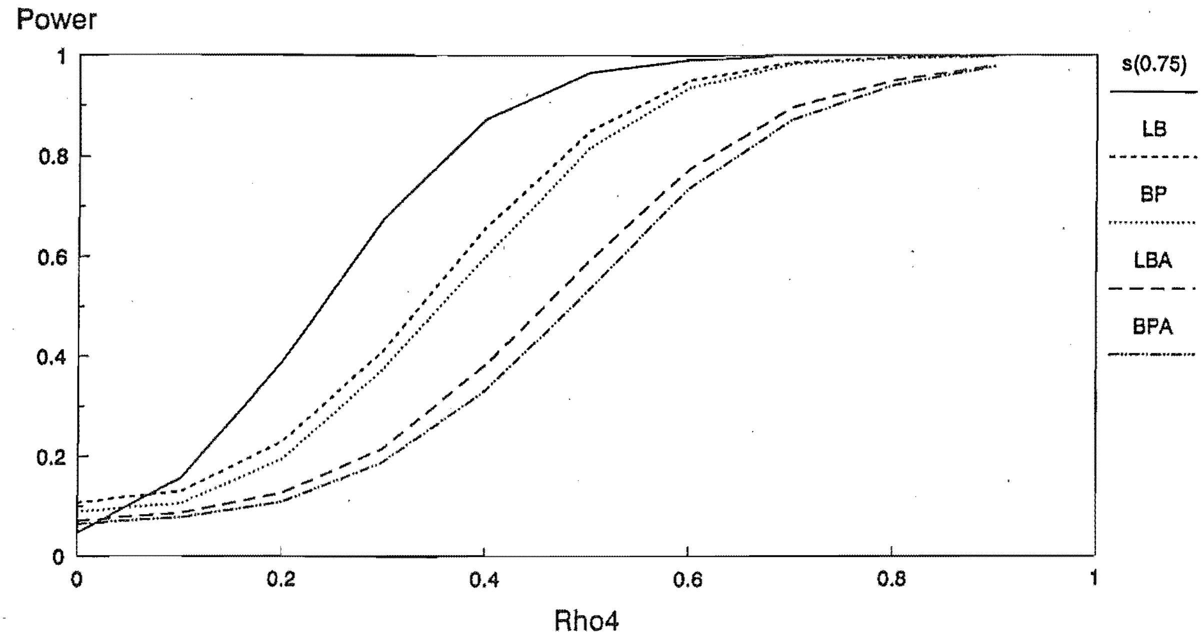


Figure 8.12a

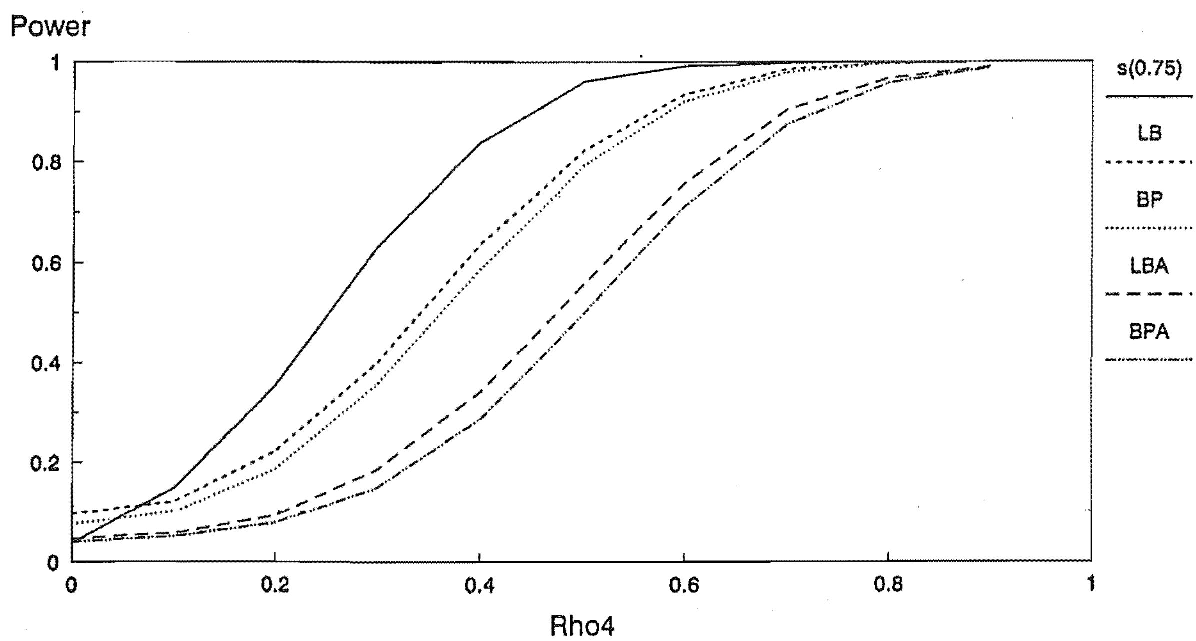
Power of Several Tests with AR(4)-ARCH(1) Errors
Data Matrix X11; Weiss Model



Alpha1=0.4

Figure 8.12b

Power of Several Tests with AR(4)-GARCH(1,1) Errors
Data Matrix X11; BHL Model



Alpha1=0.2 Beta1=0.4

CHAPTER 9

CONCLUSION

This thesis has been concerned with the properties of tests for serial independence in mis-specified linear regression models. The tests studied are all either in common usage (such as the Durbin-Watson and Ljung-Box tests) or closely related to standard tests but with theoretical advantanges which are claimed to improve their performance under certain conditions (*e.g.* the $s(\rho_1)$ point optimal tests and the "robust" LM tests).

The literature on autocorrelation testing under mis-specification was shown to be relatively sparse and incomplete in several important respects. Of the several interesting avenues of research identified in chapter 4, three were selected for particular attention in the subsequent chapters.

The new research began in chapter 5 with a study of methods by which the covariance matrix from a regression could simultaneously exhibit heteroscedasticity and autocorrelation. In the case of unconditional heteroscedasticity, it was shown that the autocorrelated disturbance, rather than its corresponding innovation term, should be assumed heteroscedastic. This allows the degrees of each process to be independently controlled but still leaves some flexibility as to the precise form of the covariance matrix. Some connections with a variance components model (such as an errors in variables framework) and a random coefficients model were established. The consideration given to the joint modelling of autocorrelation and heteroscedasticity in the (relatively recent) conditional variance literature was noted and the two proposed models were developed.

In chapter 6 a thorough study was made of the power properties of a set of exact tests against AR(1) alternatives when (unconditional) heteroscedasticity is present. Several theorems were established concerning the behaviour of power functions as the errors approach non-stationarity and, in the case of the variance components model (V*), for all non-zero AR parameter values. These theoretical results suggested that the effects of heteroscedasticity in the variance components model were likely to be more severe than in the other model considered (that identified with covariance matrix V**) throughout the parameter space. It was also shown, however, that the powers of the tests considered, in the presence of a V** covariance matrix, can approach zero as the error process approaches non-stationarity. Numerical evaluations of a comprehensive set of power functions supported these theoretical results and reinforced a general theme. There is no guarantee that the popular tests against AR(1) errors have any desirable power properties when the errors are additionally heteroscedastic.

In chapter 7 we assessed the powers of the same group of exact tests when the true error process was one of several non-AR(1) models. The three error processes selected for study were AR(2), MA(1) and a combined AR(1)-AR(4) model.

The first of these has received some attention in the literature already. It was shown that Blattberg's (1973) suggestion that the power of the DW test in such a model was related to the size of the first order autocorrelation coefficient, holds only in very special cases. We also showed that the Kadiyala (1970) based tests, which possess admirable power on the presence of very choppy data, such as Watson's (1955) matrix, have unacceptably large true sizes when used with these same data in the presence of AR(2) errors.

In the case of MA(1) errors our major conclusion was that although the ADW and DW tests are LBI and approximately LBI against such processes, this should not be used by researchers as indicating that non-rejection of serial independence on the basis of such a test gives any information about the likelihood of MA(1) errors. The reason for this is that all of the tests studied here have considerably lower power against MA(1) error processes than against AR(1) errors with the same parameter values.

The research of chapter 7 concluded with a study of the effects of ignoring a quarterly seasonal pattern in the regression residuals. Using a variety of data we showed that the additional presence of a fourth order component seriously weakens all of the tests studied. Thus, a test against an AR(1) alternative is much less likely to detect even the first order component when the true process combines both AR(1) and AR(4) errors.

Returning to the consideration of testing when heteroscedasticity is present, chapter 8 explored the effects of various GARCH processes on tests for serial independence. Using two different models which have been advanced for the joint modelling of autocorrelation and GARCH, we employed the Monte Carlo method to study the powers of the tests most relevant to such models. The tests used were drawn from the applied financial econometrics literature and included several well known asymptotic procedures as well as recently proposed "robust" versions. The thorough investigation of a large range of tests, which is reported in this chapter, has important implications for researchers in this rapidly expanding field. The asymptotic tests which are routinely used in the literature have true sizes which are substantially larger than their nominal levels. The "robust" versions of these tests have more reliable sizes but very flat power functions. Subject to the appropriate AR alternative model being formulated, we have shown conclusively that exact tests against simple AR(p) errors are remarkably robust to GARCH errors. Finally, these results appear to be invariant to certain (symmetric) conditional distributions with thicker tails than the normal distribution.

The aim of this thesis was to provide useful information about the consequences of various model mis-specifications for the powers of some autocorrelation tests. That aim has been fulfilled, but in the process several other questions have arisen which deserve further study. A brief outline of two of these is in order. First, the questions addressed in chapters 6 and 8 should be studied from the converse point of view, studying the properties of tests against heteroscedastic alternatives in the joint AR and heteroscedastic models used above. There is a small literature on these topics which is by no means complete. A comprehensive study of this topic, combined with the results presented in this thesis, may permit useful recommendations about the appropriate order in which tests for serial independence and homoscedasticity should be conducted. The second major topic which is foreshadowed by this thesis is the construction of a flexible and powerful exact procedure for the study of the serial correlation properties of regression residuals. The very poor performance of AR(1) tests in the presence of combined AR(1) and AR(4) errors suggests that an alternative method is needed, while the excellent properties of exact tests in the presence of GARCH errors underlines the benefits of exact tests. Work in progress is addressing this issue through a sequential testing procedure.

REFERENCES

Akaike, H., (1973), Information theory and an extension of the maximum likelihood principle, in B.N.Petrov and F.Csaki (eds.), 2*nd. Symposium on Information Theory*, Akademini Kiado, Budapest.

Ali, M.M. and C. Giacotto, (1984), A study of several new and existing tests for heteroscedasticity in the general linear model, *Journal of Econometrics*, 26, 355-374.

Ali, M.M. and S.C. Sharma, (1993), Robustness to non-normality of the Durbin-Watson test for autocorrelation, *Journal of Econometrics*, 57, 117-136.

Anderson, R.L., (1942), Distribution of the serial correlation coefficients, *Annals of Mathematical Statistics*, 13, 1-13.

Anderson, T.W., (1948), On the theory of testing serial correlation, *Skandinavisk Aktuarietidskrift*, 31, 88-116.

Baillie, R.T. and T. Bollerslev, (1989), The message in daily exchange rates: A conditional variance tale, *Journal of Business and Economic Statistics*, 7, 297-305.

Bartels, R and D.G. Fiebig, (1990), Integrating direct metering and conditional demand analysis for estimating end-use loads, *Energy Journal*, 1990, 79-97.

Bartels, R., and J. Goodhew, (1981), The robustness of the Durbin-Watson test, *Review of Economics and Statistics*, 63, 136-139.

Bartlett, M.S., (1946), On the theoretical specification of sampling properties of autocorrelated time series, *Journal of the Royal Statistical Society, B*, 8, 27-41.

Bera, A.K. and S. Lee, (1991), Information matrix test, Parameter heterogeneity and ARCH: a synthesis, *Review of Economic Studies*, forthcoming.

Bera, A.K., M.L.Higgins and S.Lee, (1992), Interaction between autocorrelation and conditional heteroscedasticity: a random coefficient approach, *Journal of Business and Economic Statistics*, 10, 133-142.

Bera, A.K., and M.L. Higgins, (1993), A survey of ARCH models: properties, estimation and testing, *Journal of Economic Surveys* forthcoming; also forthcoming in Oxley, L.T. *et al.* (eds.), *Surveys in Econometrics*, Blackwell, Oxford (1993).

Bera, A.K. and C.M. Jarque, (1981) Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence, *Economics Letters*, 7, 313-318.

Berenblut, I.I. and G.I. Webb, (1973), A new test for autocorrelated errors in the linear regression model, *Journal of the Royal Statistical Society, B*, 35, 33-50.

Bickel, P.J., (1978), Using residuals robustly I: tests for heteroscedasticity, non-linearity, *Annals of Statistics*, 6, 266-291.

Blattberg, R.C., (1973), Evaluation of the power of the Durbin-Watson statistic for non-first order serial correlation alternatives, *Review of Economics and Statistics*, 55, 508-515.

Bollerslev, T., (1986), Generalised autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31, 307-327.

Bollerslev, T., R.Y. Chou and K.F.Kroner, (1992), ARCH modeling in finance: A review of the theory and empirical evidence, *Journal of Econometrics*, 52, 5-59.

Box, G.E.P. and D.A. Pierce, (1970), Distribution of the residual autocorrelations in ARIMA time series models, *Journal of the American Statistical Association*, 65, 1509-1526.

Breusch, T.S., (1978), Testing for autocorrelation in dynamic linear models, *Australian Economic Papers*, 17, 334-355.

Breusch, T.S., (1979), Conflict Among Criteria For Testing Hypotheses: Extensions and Comments, *Econometrica*, 47, 203-207.

Breusch, T.S. and A.R. Pagan, (1979), A simple test for heteroscedasticity and random coefficient variation, *Econometrica*, 47, 1287-1294.

Brooks, R.D., (1992), Using a sequence of point optimal tests to select a varying coefficient model, mimeo., Department of Economics and Finance, Royal Melbourne Institute of Technology.

Cassing, S.A. and K.J. White, (1983), An examination of the eigenvector condition in the Durbin-Watson test, *Australian Journal of Statistics*, 25, 17-22.

Chesher, A.D., (1984), Testing for neglected heterogeneity, *Econometrica*, 52, 865-872.

Chou, R.Y. (1988), Volatility Persistence and Stock Valuations: Some Empirical Evidence Using GARCH, *Journal of Applied Econometrics*, 3, 279-294.

Chow, G.C., (1960), Tests of equality between sets of coefficients in two linear regressions, *Econometrica*, 28, 591-605.

Cochrane, D. and G.H. Orcutt, (1949), Application of least squares regression to relationships containing auto-correlated error terms, *Journal of the American Statistical Association*, 44, 32-61.

Cox, D.R., (1962), Some further results on separate families of hypotheses, *Journal of the Royal Statistical Society, B*, 38, 45-53.

Cox, D.R., (1970), *Analysis of Binary Data*, Methuen, London.

Crowder, M.J.,(1976), Maximum Likelihood Estimation for Dependent Observations, *Journal of the Royal Statistical Society*, B, 38, 45-53.

Cumby, R.E. and J. Huizinga, (1992), Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regression, *Econometrica*, 60, 185-196.

Davies, R.B., (1980), The distribution of a linear combination of chi square random variables, *Applied Statistics*, 29, 323-333.

Demos, A. and Sentana, E., (1991), Testing for GARCH effects: a one-sided approach, mimeo., Financial Markets Group, London School of Economics.

Dent, W.T. (1973), A power study of several tests for autocorrelation, *New Zealand Economic Papers*, 7, 109-120.

Dhrymes, P.J., (1971), *Distributed lags: Problems of estimation and formulation*, Holden-Day, San Francisco.

Dickey, D.A. and W.A. Fuller, (1981), Likelihood ratio tests for autoregressive time series with a unit root, *Econometrica*, 49, 1057-1072.

Diebold, F.X., (1986), Testing for serial correlation in the presence of ARCH, *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 323-328.

Diebold, F.X, and M, Nerlove, (1989), The dynamics of exchange rate volatility: a multivariate latent factor ARCH model, *Journal of Applied Econometrics*, 4, 1-21.

Domowitz, I. and C.S. Hakkio, (1985), Conditional variance and the risk premium in the foreign exchange market, *Journal of International Economics*, 19, 47-66.

Dubbleman, C., (1978) *Disturbances in the linear model: Estimation and hypothesis testing*, Martinus Nijhof, The Hague.

Durbin, J., (1954), Errors in variables, *Review of the International Statistical Institute*, 22, 23-32.

Durbin, J., (1980), The approximate distribution of partial serial correlation coefficients calculated from regression on Fourier series, *Biometrika*, 67, 335-349.

Durbin, J., (1970), Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables, *Econometrica*,38, 410-421.

Durbin, J. and G.S. Watson, (1950), Testing for serial correlation in least squares regression I, *Biometrika*, 37, 409-428.

Durbin, J. and G.S. Watson, (1951), Testing for serial correlation in least squares regression II, *Biometrika*, 38, 159-178.

Durbin, J. and G.S. Watson, (1971), Testing for serial correlation in least squares regression III, *Biometrika*, 58, 1-19.

Easton, P.D. and M.E. Zmijewski, (1989), Cross-sectional variation in the stock market response to accounting earnings announcements, *Journal of Accounting and Economics*, 11, 117-141.

Engle, R.F.,(1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987-1008.

Engle, R.F.,(1984), Wald Likelihood Ratio and Lagrange Multiplier Tests in Econometrics, Chapter 13 in *Handbook of Econometrics, Vol* II, Z. Griliches, and M.D. Intriligator, (eds.), North-Holland, Amsterdam.

Engle, R.F., D.F. Hendry and D. Trumble, (1985), Small sample properties of ARCH estimators and tests, *Canadian Journal of Economics*, 18, 66-93.

Epps, T.W. and M.L. Epps, (1977), The robustness of some standard tests for autocorrelation and heteroscedasticity when both problems are present, *Econometrica*, 45, 745-753.

Evans, G.B.A. and N.E. Savin, (1982), Conflict Among the Criteria Revisited; The W, LR and LM Tests, *Econometrica*, 50, 737-748.

Evans, M.A., (1983), Inference and non-ideal conditions in the linear regression model, unpublished Ph.D. thesis, Monash University.

Evans, M.A., (1992), Robustness and size of tests of autocorrelation and heteroscedasticity to nonnormality, *Journal of Econometrics*, 51, 7-24.

Evans, M.A. and M.L. King, (1985a), Critical value approximations for tests of linear regression disturbances, *Australian Journal of Statistics*, 27, 68-83.

Evans, M.A. and M.L. King, (1985b), A Point Optimal Test for Heteroscedastic Disturbances, *Journal of Econometrics*, 27, 163-178.

Evans, M.A., and M.L. King, (1988), A further class of tests for heteroscedasticity, *Journal of Econometrics*, 37, 265-276.

Faust, J., (1992), When are variance ratio tests for serial dependence optimal?, *Econometrica*, 60, 1215-1226.

Fraser, D.A.S., I. Guttman and G.P.H. Styan (1976), Serial correlation and distributions on the sphere, *Communications in Statistics*, A5, 97-118.

Gastwirth, J.L. and M.R.Selwyn, (1980), The robustness properties of two tests for serial correlation, *Journal of the American Statistical Association*, 75, 138-141.

Geary, R.C., (1970), Relative efficiency of count of sign changes for assessing residual autoregression in least squares regression, *Biometrika*, 57, 123-127.

Geweke, J., (1988), Diagnostics for the diagnostics, *Econometric Reviews*, 7, 59-62.

Giles, D.E.A. and M. Beattie, (1987), Autocorrelation pre-test estimation in models with a lagged dependent variable. In M.L. King and D.E.A. Giles (eds.), *Specification Analysis in the Linear Model* (pp. 99-116). Routledge & Kegan Paul, London.

Giles, D.E.A., J.A. Giles and J.K. Wong, (1992), Testing for ARCH-GARCH errors in a mis-specified regression, Discussion Paper 9201, Department of Economics, University of Canterbury.

Giles, D.E.A., and O. Lieberman, (1992), Some properties of the Durbin-Watson test after a preliminary t-test, *Journal of Statistical Computation and Simulation*, 41, 219-227.

Giles, D.E.A. and J.P. Small, (1991), The power of the Durbin Watson test when the errors are heteroscedastic, *Economics Letters*, 36, 37-41.

Giles, J.A. and D.E.A. Giles, (1993) Pre-test estimation and testing in econometrics: Recent developments, forthcoming in *Journal of Economic Surveys;* also forthcoming in L.T. Oxley *etal.* (eds.), *Surveys in Econometrics*, Blackwell, Oxford (1993).

Glejser, H., (1969), A new test for heteroscedasticity, *Journal of the American Statistical Association*, 64, 316-323.

Godfrey, L.G., (1978), Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables, *Econometrica*, 46, 1293-1302.

Goldberg, S., (1958), *Difference Equations*, John Wiley, New York.

Goldfeld, S.M. and R.E. Quandt, (1965), Some tests for homoscedasticity, *Journal of the American Statistical Association*, 60, 539-547.

Goldfeld, S.M. and R.E. Quandt, (1972), *Nonlinear methods in econometrics*, North Holland, Amsterdam.

Golub, G.H., (1965), Numerical methods for solving linear least squares problems, *Numerische Mathematik*, 7, 207-216.

Gregory, A.W. and M.R. Veall, (1985), Formulating Wald Tests of Nonlinear Restrictions, *Econometrica*, 53, 1465-1468.

Hájek, J. and Z. Šidák, (1967), *The theory of rank tests*, Academic Press, New York.

Harrison, M.J. and B.P.M. McCabe (1975), Autocorrelation with heteroscedasticity: A note on the robustness of the Durbin-Watson, Geary and Henshaw tests, *Biometrika*, 62, 214-215.

Harrison, M.J. and B.P.M. McCabe (1979), A new test for heteroscedasticity based on ordinary least squares residuals, *Journal of the American Statistical Association*, 74, 494-499.

Harvey, A.C., (1976), Estimating regression models with multiplicative heteroscedasticity, *Econometrica*, 44, 461-465.

Harvey, A.C., (1981), *Time series models*, Philip Allan and Humanities Press, London.

Harvey, A.C. and G.D.A. Phillips, (1974), A comparison of the power of some tests for heteroscedasticity in the general linear model, *Journal of Econometrics*, 2, 307-316.

Harvey, A.C. and G.D.A. Phillips, (1980), Testing for serial correlation in simultaneous equations models, *Econometrica*, 48, 747-759.

Hendry, D.F., (1984), Monte Carlo Experimentation in Econometrics, Chapter 16 of *Handbook of Econometrics*, Vol II, Z.Griliches, and M.D.Intriligator, (eds,), North-Holland, Amsterdam.

Henshaw, R.C., (1966), Testing single equation least squares regression models for autocorrelated disturbances, *Econometrica*, 34, 646-660.

Hildreth, C. and J.P. Houck, (1968), Some estimators for a linear model with random coefficients, *Journal of the American Statistical Association*, 70, 872-879.

Hoque, A., (1991), An application and test for a random coefficient model in Bangladesh agriculture, *Journal of Applied Econometrics*, 6, 77-90.

Hsiao, C., (1986), *Analysis of Panel Data*, Cambridge University Press, Cambridge.

Hsieh, D.A., (1989), Modeling Heteroscedasticity in Daily Foreign Exchange Rates, *Journal of Business and Economic Statistics*, 7, 307-317.

Inder, B.A., (1985), Finite sample power of tests for autocorrelation in models containing lagged dependent variables, *Economics Letters*, 14, 179-185.

Inder, B.A., (1986), An approximation to the null distribution of the Durbin-Watson ststistic in models containing lagged dependent variables, *Econometric Theory*, 2, 413-428.

Judge, G.G., W.E. Griffiths, R.C. Hill and T-C. Lee, (1980), *The theory and practice of econometrics*, Wiley, New York.

Kadiyala, K.R., (1970), Testing for the independence of regression disturbances, *Econometrica*, 38, 97-117.

Kendall, M.G., (1976), *Time-Series*, 2nd. ed., Griffin, London.

Kenkel, J.L., (1974), Some small sample properties of Durbin's tests for serial correlation in regression models containing lagged dependent variables, *Econometrica*, 42, 763-769.

Kenkel, J.L., (1975), Small sample tests for serial correlation in models containing lagged dependent variables, *Review of Economics and Statistics*, 57, 383-386.

King, M.L., (1979), Some aspects of statistical inference in the linear regression model, unpublished Ph.D. thesis, University of Canterbury.

King, M.L., (1980), Robust tests for spherical symmetry and their application to least squares regression, *The Annals of Statistics*, 8, 1286-1271.

King, M.L., (1981), The alternative Durbin Watson test: an assessment of Durbin and Watson's choice of test statistic, *Journal of Econometrics*, 17, 51-66.

King, M.L., (1982), Testing for a serially correlated component in regression disturbances, *International Economic Review*, 23, 577-582.

King, M.L., (1983), Testing for moving average regression disturbances, *Australian Journal of Statistics*, 25, 23-34.

King, M.L., (1985), A point optimal test for autoregressive disturbances, *Journal of Econometrics*, 27, 21-37.

King, M.L. and M.A. Evans, (1986), Testing for block effects in regression models based on survey data, *Journal of the American Statistical Association*, 81, 677-679.

King, M.L. and D.E.A. Giles, (1984), Autocorrelation pre-testing in the linear model: Estimation, testing and prediction, *Journal of Econometrics*, 25, 35-48.

Knight, J.L., (1983), The sensitivity of the Durbin-Watson test to the assumption of normal disturbances, Paper presented to the European Meetings of the Econometric Society, Pisa.

Koerts, J. and A.P.J. Abrahamse, (1969), *On the theory and application of the general linear model*, Rotterdam University Press, Rotterdam.

Koopmans, T.C. (1942), Serial correlation and quadratic forms in normal variables, *Annals of Mathematical Statistics*, 13, 14-33.

Kramer, G., (1971), On the Durbin-Watson bounds test in the case of regression through the origin, *Jahrbucher fur Nationalokonomie und Statistik*, 185, 345-358.

Krämer, W., (1985), The power of the Durbin-Watson test for regressions without an intercept, *Journal of Econometrics*, 28, 363-370.

Krämer, W. and H. Zeisel, (1990), Finite sample power of linear regression autocorrelation tests, *Journal of Econometrics*, 43, 363-372.

Lafontaine, F. and K.J. White, (1986), Obtaining any Wald statistic you want, *Economics Letters*, 21, 35-40.

Lee, J.H.H., (1991), A Lagrange multiplier test for GARCH models, *Economics Letters*, 37, 265-271.

Lee, J.H.H. and M.L. King, (1991), A locally most mean powerful based score test for ARCH and GARCH regression disturbances, Working Paper No. 9/91, Department of Econometrics, Monash University,

Lehmann, E.L.,(1986) Testing Statistical Hypotheses, 2nd ed., John Wiley & Sons, New York.

Lehmann, E.L. and S. Stein, (1948), Most powerful tests of composite hypotheses, I, normal distributions, *Annals of Mathematical Statistics*, 19, 495-516.

L'Esperence, W.L. and D. Taylor, (1975), The power of four tests of autocorrelation in the linear regression model, *Journal of Econometrics*, 3, 1-21.

Ljung, G.M. and G.E.P. Box, (1978), On a measure of lack of fit in time series models, *Biometrika*, 65, 297-303.

Liu, L.M. and G.C. Tiao, (1980), Random coefficient first order autoregressive models, *Journal of Econometrics*, 13, 305-325.

Mandelbrot, B., (1963), The variation of certain speculative prices, *Journal of Business*, 36, 394-419.

Milhöj, A., (1985); The moment structure of ARCH processes, *Scandanavian Journal of Statistics*, 12, 281-292.

Nelson, D.B., (1991), Conditional heteroscedasticity in asset return: a new approach, *Econometrica*, 59, 347-370.

Nelson, D.B. and C.Q. Cao, (1992), Inequality constraints in the univariate GARCH model, *Journal of Business and Economic Statistics*, 10, 229-235.

Nerlove, M., and K.F. Wallis, (1966), Use of the Durbin-Watson statistic in inappropriate situations, *Econometrica*, 34, 235-238.

Neymann, J. and E.S. Pearson, (1933), *Philosophical Transactions of the Royal Society of London, A*, 231, 289-337.

Newey, W.K., and K.D. West, (1987), Hypothesis Testing with Efficient Method of Moments Estimation, *International Economic Review*, 28, pp 777-87.

Orcutt, G.H., and D. Cochrane, (1949), A sampling study of the merits of autoregressive and reduced form transformations in regression analysis, *Journal of the American Statistical Association*, 44, 365-372.

Pagan, A., (1973), Efficient estimation of models with composite disturbance terms, *Journal of Econometrics*, 1, 329-340.

Park, S. (1975), On the small sample power of Durbin's *h* test, *Journal of the American Statistical Association*, 70, 60-63.

Pesaran, H., (1974), On the general problem of model selection, *Review of Economic Studies*, 41, 153-171.

Pereira, B. De B.,(1977) Discriminating Among Separate Models: A Bibliography, *International Statistical Review*, 45, pp 163-72.

Phillips, G.D.A. and A.C. Harvey, (1974), A simple test for serial correlation in regression analysis, *Journal of the American Statistical Association*, 69, 939-939.

Phillips, P.C.B. and J.Y. Park, (1988), On the formulation of Wald tests of nonlinear restrictions, *Econometrica*, 56, 1065-1083.

Rao, C.R.,(1973), *Linear Statistical Inference and its Applications*, Wiley, New York.

Revankar, N.S., (1980), Analysis of regressions containing serially correlated and serially uncorrelated error components, *International Economic Review*, 21, 185-199.

Savin, N.E., (1976), Conflicts Among Testing Procedures in a Linear Regression Model with Autoregressive Disturbances, *Econometrica*, 44, 1303-1313.

Schmidt. P., and D.K. Guilkey, (1975), Some further evidence on the power of the Durbin-Watson and Geary tests, *Review of Economics and Statistics*, 57, 379-382.

SHAZAM, (1993), *Econometrics Computer Program, Users Reference Manual, Version* 7, McGraw-Hill, New York.

Silvey, S.D.,(1959), The Lagrange Multiplier Test, *Annals of Mathematical Statistics*, 30, 387-407.

Small, J.P., (1993), The limiting power of point optimal autocorrelation tests, forthcoming in *Communications in Statistics, Theory and Methods*.

Small, J.P., D.E.A. Giles and K.J. White, (1993), The exact powers of some autocorrelation tests when relevant regressors are omitted, Discussion Paper 9305, Department of Economics, University of Canterbury.

Smith, V.K., (1976), The estimated power of several tests for autocorrelation with non-first-order alternatives, *Journal of the American Statistical Association*, 71, 879-883.

Spencer, B.G., (1975), The small sample bias of Durbin's tests for serial correlation when one of the regressors is the lagged dependent variable and the null hypothesis is true, *Journal of Econometrics*, 3, 249-254.

Szröeter, J. (1978), A class of parametric tests for heteroscedasticity in linear econometric models, *Econometrica*, 46, 1311-1328.

Theil, H., (1965), The analysis of disturbances in regression analysis, *Journal of the American Statistical Association*, 60, 1067-1079.

Theil, H, (1971), *Principles of Econometrics*, Wiley, New York.

Tiao, G.C., and I. Guttman, (1967), Analysis of outliers with adjusted residuals, *Technometrics*, 9, 541-559.

Tsay, R.S., (1987) Conditional heteroscedastic time series models, *Journal of the American Statistical Association*, 82, 590-604.

Vinod, H.D., (1973), Generalisation of the Durbin-Watson statistic for higher order autoregressive processes, *Communications in Statistics*, *B*, 2, 115-144.

Wallis, K.F., (1972), Testing for fourth order autocorrelation in quarterly regression equations, *Econometrica*, 40, 617-636.

Ward, L.L., (1973), Is uncorrelating the residuals worth it?, Unpublished M.Sc. thesis, McGill University.

Watson, G.S., (1955), Serial correlation in regression Analysis I, *Biometrika*, 42, 327-341.

Webb, G.I., (1973), Autocorrelations and the general linear model, unpublished Ph.D. thesis, The City University, London.

Weber, J.E. and D.E. Monarchi, (1982), Performance of the Durbin-Watson test and WLS estimation when the disturbance term includes serial dependence in addition to first order autocorrelation, *Journal of the American Statistical Association*, 77, 117-128.

Weiss, A.A., (1984), ARMA models with ARCH errors, *Journal of Time Series Analysis*, 5, 129-143.

White, H., (1980), A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica*, 48, 817-838.

White, H., (1982), Maximum likelihood estimation of misspecified models, *Econometrica*, 50, 1-25.

Wolff, C.C.P., (1988), Autoregressive conditional heteroscedasticity; A comparison of ARCH and random coefficient models, *Economics Letters*, 27, 141-143.

Wooldridge, J.M., (1991), On the application of robust regression based diagnostics to models of conditional means and conditional variances, *Journal of Econometrics*, 47, 5-46.

Zeisel, H., (1989), On the power of the Durbin-Watson test under high autocorrelation, *Communications in Statistics, Theory and Methods*, 18, 3907-3916.

Zellner, A., (1992), Statistics, Science and Public Policy, *Journal of the American Statistical Association*, 87, 1-6.